Heterogeneous Intermediary Asset Pricing

Mahyar Kargar*

September 20, 2019
First draft: May 5, 2018

Abstract

I show that the composition of the financial sector has important asset pricing implications beyond the health of the aggregate financial sector. To assess the impact of massive balance sheet adjustments within the intermediary sector during the Great Recession and resolve conflicting asset pricing evidence, I propose a dynamic asset pricing model with heterogeneous intermediaries facing financial frictions. Asset flows between intermediaries are quantitatively important for both level and variation of risk premia. An empirical measure of the composition of the intermediary sector negatively forecasts future excess returns and is priced in the cross-section with a positive price of risk.

Keywords: Heterogeneous Intermediaries, Intermediary Asset Pricing, Predictability, Leverage Cyclicality.

JEL Classification: G12, G11, G21.

*Gies College of Business, University of Illinois Urbana-Champaign, 435 Wohlers Hall, 1206 S. Sixth Street, Champaign, IL 61820, phone: (217) 300-7640, e-mail: kargar@illinois.edu. I am incredibly grateful to Andrea Eisfeldt and Mike Chernov, as well as Tyler Muir, William Mann, and Pierre-Olivier Weill for invaluable support and guidance. I especially thank Alexi Savov for valuable discussions and feedback in the early stages of this project. For helpful comments and discussions I thank Darren Aiello, Andy Atkeson, Tony Bernardo, Bruce Carlin, Murray Carlson, Alex Fabisiak, Salil Gadgil, Chady Gemayel, Valentin Haddad, Barney Hartman-Glaser, Lars Lochstoer, Stavros Panageas, Rob Richmond, Bill Schiesser, Andrés Schneider, Avanidhar Subrahmanyam, Clinton Tepper, Ivo Welch, Geoffery Zheng, Irina Zviadadze, and participants at the 2019 NFA conference in Vancouver, University of Colorado Leeds, University of Illinois Gies, Indiana University Kelley, OSU Fisher, Rice University Jones, UCLA Anderson, and UCLA Economics seminars.

1 Introduction

The financial sector witnessed a massive restructuring of its balance sheet during the Great Recession: over the period from the first quarter of 2008 to the fourth quarter of 2009, which includes the most dramatic episode of the crisis in the fall of 2008, (i) broker-dealers drastically reduced asset holdings by approximately \$1.7 trillion (a 35% drop) while commercial banks increased total asset by nearly \$1 trillion (a 7.5% rise), and (ii) broker-dealers reduced leverage by about 47% while bank holding companies increased leverage by approximately 72%. This evidence is at odds with canonical intermediary asset pricing models which feature a *representative* financial sector. This paper makes two main contributions. First, I extend the existing frameworks to explain these massive asset flows. Second, I apply my model to study its empirical asset pricing implications for time-series predictability and the cross-section of assets.

To explain these large asset flows and resolve the puzzling evidence presented above, I present a parsimonious dynamic asset pricing model with two key features: (i) intermediaries *heterogeneous* in their risk-bearing capacity, and (ii) state-dependent margin constraints. I show that the *composition* of the intermediary sector has important asset pricing implications beyond the health of the overall financial sector previously considered in the literature. I quantify the importance of this heterogeneity for the level and variation of the risk premium.

Guided by my model, I present two main empirical results that transcend the specific 2008 crisis episode. First, the wealth share of broker-dealers in the financial sector, a measure of the composition of the intermediary sector, strongly forecasts future market excess returns with additional predictive power beyond many popular forecasting variables in the literature. Second, this measure of heterogeneity has strong explanatory power for the cross-section of asset: shocks to the relative wealth share of broker-dealers in the financial sector, explain the cross-section of equity and bond returns about as well or better than existing intermediary asset pricing models.

As a corollary, my model reconciles seemingly contradictory asset pricing evidence from recent empirical evaluations of *representative* intermediary-based models. In particular, Adrian, Etula, and Muir (2014a) (henceforth, AEM) and He, Kelly, and Manela (2017) (henceforth, HKM) find opposite signs for the price of intermediary leverage shocks in the cross-section of assets (positive and negative, respectively). Importantly, AEM and HKM measure intermediary leverage in *different* parts of the financial sector: security broker-dealers, and bank holding companies, respectively. The economic mechanism of my model, presented below, implies opposite leverage dynamics for different parts of the financial sector, resolving this puzzling evidence.

The model features two main ingredients. First, I assume agents differ in their attitudes toward risk: two financial intermediaries (labeled A and B) and a household sector (C agents) in order of increasing risk aversion. I think of A and B intermediaries as broker-dealers and banks, respectively. This is consistent with the evidence that more aggressive hedge fund and broker-dealers have higher leverage on their balance sheets compared to more passive commercial banks. In equilibrium, both intermediaries hold levered positions in the risky asset financed by borrowing from more risk averse agents. Second, investors face financial frictions in the form of *occasionally binding* state-dependent margin constraints.[1]

While I extend a heterogeneous-agent model with occasionally binding leverage constraints to a setting with three agents (households and two levered intermediaries) and recursive preferences, the main contributions of this work are: (i) presenting a dynamic framework to quantitatively analyze the importance of heterogeneity in the financial sector, (ii) showing its ability to match patterns of heterogeneity observed in the data, and, (ii) empirically validating its asset pricing predictions.

Although all agents face margin constraints, in equilibrium, only more aggressive A type intermediaries face constraints that occasionally bind.[2] This calibration assumption has empirical support in the data: hedge funds and broker-dealers primarily rely on collateralized repo financing with haircuts, while the commercial banking sector has access to more stable funding sources, such as insured deposits and discount window lending from a central bank.[3] This is consistent with findings of Gorton and Metrick (2012) that repo haircuts tend to rise during financial crises leading

[1] In the dynamic model, the margin constraint does not bind in non-distress states and binds only after a sequence of sufficiently negative aggregate shocks.

[2] In the baseline calibration of the model, the margin constraint does not bind for moderately risk-averse B intermediaries (banks) in equilibrium and only *occasionally* binds for A-types (dealers). This assumption can be relaxed by allowing the margin constraint to bind for B agents as well without affecting the main results of the model. In a very severe financial crisis where *both* intermediaries face a binding leverage constraint, the C agents (households) need to hold the excess supply of the risky asset that intermediaries have to sell to reduce leverage.

[3] In the second quarter of 2018, approximately 36% of total financial assets and 50% of total liabilities for security broker-dealers are due to lending and borrowing in the repo market, respectively. For private depository institutions, approximately 73% of total liabilities are comprised of checkable, time and savings deposits. Source: Financial Accounts of the United States (Flow of Funds).

to binding margin constraint for broker-dealers. Begenau, Bigio, Majerovitz, and Vieyra (2019) provide cross-sectional evidence suggesting that neither regulatory nor market leverage constraints bind strictly for most banks.

My model features occasionally binding, time-varying margin constraints: the level of margin required in the model is state-dependent; and (inversely) linked to endogenously determined return volatility resembling an approximate Value-at-Risk (VaR) rule. Since there is empirical evidence that return volatility is higher in bad times (Schwert, 1989, for example), such an approach is consistent with findings of Gorton and Metrick (2012) that haircuts tend to rise in financial crises tightening the constraint.

The primary economic mechanism of the model is as follows: following a negative shock, more aggressive A type intermediaries face tighter and eventually binding margin constraints and are forced to reduce leverage by selling assets. To clear the risky asset market, the less aggressive B intermediaries have to take on a larger portion of the asset than they would in the absence of constraints. The risk premium must increase to compensate them for bearing more risk. Since constraints are more likely to bind in high marginal utility states where intermediary wealth shares are low, consistent with empirical evidence, the model generates opposite cyclical dynamics for leverage of different intermediaries: procyclical for more aggressive and countercyclical for less aggressive intermediaries.[4]

Heterogeneity in intermediaries' risk appetite and margin constraints are *both* necessary to match and understand asset reallocations within the financial sector that are consistent with observed patterns. In a model with *representative* intermediaries, only the aggregate wealth share of the financial sector matters for asset prices. Thus, in the absence of heterogeneity among intermediaries, models are unable to match asset flows within the financial sector. Similarly, in a setting with *no financial frictions*, an adverse shock reduces intermediaries' risk-bearing capacity and results in a fall in prices and a direct increase in leverage for *both* regardless of their degree of risk tolerance. Therefore, without margin constraints, both intermediaries would counterfactually lever up and down at the same time, and the model would be inconsistent with empirical evidence

[4]As mentioned above, one can think of aggressive and passive intermediaries as broker-dealers (BDs) and banks, respectively, consistent with the observation that BDs are more likely to face binding borrowing constraints in bad times than banks.

on balance sheet adjustments in the financial sector.

I show that the model's equilibrium dynamics can be described by two endogenous state variables: (i) total wealth share of the financial sector (i.e. A and B agents) in the economy, and (ii) the more aggressive intermediary's (A types) wealth share as a fraction of the total financial sector.[5] The former is the main state variable in many recent representative intermediary-based models (e.g., He and Krishnamurthy, 2012, 2013, and Brunnermeier and Sannikov, 2014). The unique feature of the model is the second state variable: it emphasizes that the composition of the financial sector is a key factor in determining asset prices. Models with a representative financial sector are silent about the composition of intermediaries and its asset pricing implications.

Risk tolerant intermediaries (A and B agents) have levered balance sheets by borrowing from more risk averse households. Leverage increases intermediaries' exposure to aggregate shocks: positive shocks result in their wealth share to increase. Following a negative aggregate shock, the wealth of levered intermediaries falls faster than that of households, and hence their share of total wealth, the first state variable in the model, declines. Moreover, since more risk tolerant intermediaries (i.e., A agents) have higher leverage than more risk averse ones, they are more likely to face binding margin constraints in bad times when they become tighter. As such, their wealth share in the financial sector, the second state variable of my model, declines as well. A key takeaway from model's economic mechanism is that both state variables exhibit *procyclical* dynamics: in high marginal utility states, wealth share of the aggregate financial sector and more aggressive intermediary's net worth share in the financial sector will *both* be low.

I then examine two quantitative implications of my model. First, I show that the composition of the financial sector is responsible for a significant fraction of risk premia variation beyond the wealth share of the aggregate financial sector. With the model's independent and identically distributed (i.i.d.) aggregate endowment, variation in risk premia is only due to the aggregate wealth share as well as the composition of the financial sector, model's two state variables. I show that

[5] This representation is not the only way one can construct the state variables in the model. I argue my construction is a natural and valid way to do so: measuring the wealth share of the financial sector in the economy (which I denote as state variable x) and that of A intermediaries relative to the financial sector (denoted by y). In many representative intermediary-based models, x is the main state variable. Moreover, with the choice of y presented here, in the absence of heterogeneity, my model collapses to one with x as the only state variable similar to the models with a representative financial sector.

approximately 20% of the variation in risk premia can be attributed to the wealth distribution among intermediaries, which is a measure of the composition of the financial sector. Thus, failing to account for heterogeneity among intermediaries can lead to missing a substantial portion of risk premia variation. This result implies a novel empirical prediction of the model: the composition of the financial sector should strongly forecast future excess returns. I later document that this is indeed the case, empirically.

Second, I use the model to quantify the asset pricing implications of massive financial flows between intermediaries observed in the 2008 crisis. As mentioned above, during 2008–2009, broker-dealers drastically reduced asset holdings and leverage (by $1.7 trillion and 47%, respectively), while banks increased both (by $1 trillion and 72%, respectively). My model implies that a dealer deleveraging episode comparable in magnitude to the one observed during the crisis leads to an approximately 55% increase in the risk premia and a 5% increase in volatility. These balance sheet adjustments have no impact on asset prices in existing models with representative intermediaries.

Next, I study the empirical implications of the model for time-series predictability and the cross-section of assets. I show that in addition to wealth share of the aggregate financial sector, keeping track of the composition of intermediaries is also crucial for determining risk premia. I define an empirical proxy for measuring this composition: the ratio of the equity of security broker-dealers to sum of equities of broker-dealers and commercial banks from the Financial Accounts of the United States (Flow of Funds).

I emphasize that A agents in the model are not meant only to represent broker-dealers (BDs). BDs are good proxies for a set of intermediaries that face tightening leverage constraints in bad times (e.g., hedge funds) whose actions resemble these agents in the model. Moreover, unlike hedge funds, broker-dealers are intermediaries for which I have access to a relatively long time-series of aggregate balance sheet data. Similarly, B agents are meant to describe intermediaries who face equity capital constraints. Commercial banks are good proxies for these intermediaries: for them, equity is hard to raise, but given their access to more stable funding sources, they can attract deposits even during financial crises.

It is also important to point out that many major broker-dealers (such as primary dealers) are subsidiaries of large bank holding companies. If the internal capital markets represent an important funding source for these dealer subsidiaries, it may seem unnecessary to treat them separately form BHCs.[6] For instance, if the commercial banking affiliate of a BHC experiences large deposit inflows during a crisis (as documented in Gatev and Strahan (2006), for example), the access to internal capital markets may substantially mitigate the financial distress in the broker-dealer subsidiary.

Federal Reserve's Regulation W prohibits BHCs from freely using deposits in their commercial banking arm to fund the broker-dealer subsidiary when repo markets collapse.[7] Moreover, Gupta (2018) documents that the internal capital markets are not frictionless and do not behave very differently from the external ones during the financial crisis. He shows that although the aggregate size of the dealer internal capital markets quadrupled from $300 billion in 2001 to $1.2 trillion by 2007, these inter-affiliate repo and securities loans collapsed by 37% in 2008.

Consistent with my model's prediction, the composition of the financial sector strongly and *negatively* predicts future excess return. Model's second state variable, which captures this wealth distribution in the financial sector, exhibits procyclical dynamics: times when broker-dealers are relatively more impaired in the financial sector, coincide with high marginal utility states where prices are low and future expected returns are high. I show that this measure of the composition of the financial sector is a strong predictor of future excess returns leading to additional predictive power beyond many established return forecasting variables in the literature.

Moreover, *shocks* to the composition of intermediaries, which I denote the heterogeneous intermediary factor (HIFac), are priced in the cross-section of asset returns with a *positive* price of risk: the HIFac alone exhibits strong explanatory power for the cross-section of equity and bond returns about as well or better than existing intermediary asset pricing factors in AEM and HKM. I further document that including aggregate intermediary leverage as a second asset pricing factor increases cross-sectional fit by at least ten percentages points, depending on whether the factor is

[6]For large BHCs, BD subsidiaries hold, on average, between 15 to 20% of total assets. For example, from FOCUS reports and 10K filings to the SEC, in 2018, total assets of J.P. Morgan Securities LLC (the BD subsidiary of J.P. Morgan Chase & Co, the BHC) represented approximately 17.6% of BHC's total assets. Similarly, Citigroup Global Markets Inc. (the BD arm) accounted for 15.6% of Citigroup's total assets in 2018.

[7]Sections 23A and 23B of Federal Reserve's Regulation W set limits on the amount of covered transactions between a bank and its affiliates and require these transactions to be collateralized and on market terms and conditions. Covered transactions include loans and other extensions of credit to an affiliate, investments in the securities of a subsidiary, purchases of assets from an affiliate, and certain other transactions that expose the bank to the risks of its affiliates. See Appendix D for more details on Regulation W.

leverage shocks for broker-dealers or bank holding companies.

In addition, I reconcile seemingly contradictory evidence in AEM and HKM by proposing a unifying general equilibrium framework with heterogeneous intermediaries. Pricing kernels in AEM and HKM measure marginal utilities of *different* financial intermediaries: broker-dealers and bank holding companies, respectively. Given the economic mechanism presented above, different parts of the financial sector exhibit opposite leverage dynamics. Therefore, it does not seem surprising that the literature with a representative financial sector arrives at conflicting asset pricing results.

I provide further evidence that heterogeneity in the financial sector is an important risk factor. Stock portfolios sorted on their exposure to HIFac (shocks to dealers' wealth share in the financial sector) exhibit monotonically increasing excess returns: the highest-beta quintile has approximately 5% higher annualized excess return relative to the lowest-beta portfolio. Existing representative intermediary asset pricing models are unable to capture these results.

I finally construct a mimicking portfolio for the heterogeneous intermediary factor from my model. Mimicking portfolios for representative intermediary factors in AEM and HKM are unable to fully span the heterogeneous intermediary factor-mimicking portfolio (FMP): I find large and highly significant alphas when I regress my model's FMP on AEM's and HKM's FMPs both individually and in bivariate regressions. This result corroborates my earlier findings: the composition of the financial sector is an important source of risk and has pricing information beyond representative intermediary asset pricing factors.

Related Literature

My paper extends macroeconomic models with a financial sector (e.g., He and Krishnamurthy, 2012, 2013, Brunnermeier and Sannikov, 2014, and Gertler and Kiyotaki, 2010) to a framework with heterogeneous intermediaries. This literature builds on financial accelerator models of Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Bernanke, Gertler, and Gilchrist (1999) which emphasize the importance of financial frictions and leverage for persistence and amplification of aggregate shocks.[8] The literature traditionally modeled intermediaries as one representative sector. Such an approach does not allow for the heterogeneity in financial intermediaries, documented in

[8]See Brunnermeier, Eisenbach, and Sannikov (2012) for a survey of macro-based models with financial frictions.

the data, to play a role in equilibrium.[9]

A few recent papers focus on the importance of a heterogeneous financial sector. The paper closest to this work is Ma (2017). In independent, contemporaneous work, he shows that an SDF estimated from a model with intermediaries heterogeneous in the tightness of their constraints exhibits higher cross-sectional R^2 than AEM and HKM factors. In contrast to this paper where I show asset reallocations within the financial sector are quantitatively important for both level and variation of risk premia, Ma (2017) focuses exclusively on explaining the cross-sectional variation of asset returns. Moreover, unlike his model where financing constraints are always binding, I study occasionally binding state-dependent leverage constraints.[10] I also fully characterize the whole dynamic system instead of merely a log-linearized representation around the steady-state in Ma (2017). Coimbra and Rey (2017) develop a model with intermediaries heterogeneous in their Value-at-Risk constraints and limited liability resulting in risk-shifting. Gertler, Kiyotaki, and Prestipino (2016) extend Gertler and Kiyotaki (2010)'s framework by incorporating a shadow banking sector alongside retail banks and allow the possibility of runs. The latter papers do not study the asset pricing implications of heterogeneous intermediaries. I contribute to this literature by showing that the composition of the financial sector has strong predictive power for excess returns of many assets and it is also priced in the cross-section of equity and bond returns.

This paper also contributes to the recent empirical intermediary asset pricing literature. As noted above, two recent papers, Adrian et al. (2014a) and He et al. (2017), evaluate the explanatory power of models where *representative* intermediaries face, respectively, debt and equity constraints for cross-sectional variations in expected returns. They find opposite signs for the estimated price of risk (and thus conflicting cyclical dynamics) for intermediary leverage. I reconcile these seemingly contradictory evidence in a unifying general equilibrium framework where the financial sector is modeled as two sectors heterogeneous in risk-bearing capacity facing margin constraints.

Finally, this work relates to the extensive literature on asset pricing implications of investor

[9]A few other recent models with explicit roles for financial intermediaries include Danielsson, Shin, and Zigrand (2012), Adrian and Shin (2014), Adrian and Boyarchenko (2015), Moreira and Savov (2017), and Drechsler, Savov, and Schnabl (2018).

[10]Although the presence of occasionally binding financial constraints makes solving the model considerably more challenging, as discussed in He and Krishnamurthy (2019), this is a necessary step to study systemic risk because financial crises are rare, and in most cases, we are interested in understanding the transitional dynamics of the economy from non-crisis states into crisis states.

heterogeneity and portfolio constraints. Dumas (1989), Wang (1996), Bhamra and Uppal (2009), Bhamra and Uppal (2009), Longstaff and Wang (2012), Gârleanu and Panageas (2015), and Santos and Veronesi (2016) study equilibrium in *frictionless* economies with two heterogeneous agents and different preference assumptions. Basak and Cuoco (1998), Kogan, Makarov, and Uppal (2007), Gârleanu and Pedersen (2011), Danielsson et al. (2012), He and Krishnamurthy (2012, 2013), Chabakauri (2013), and Rytchkov (2014) examine asset pricing implications of exogenous and endogenous portfolio constraints in economies populated by two heterogeneous agents with one or many assets where constrained agent have logarithmic or CRRA preferences. I use this general framework to study the asset pricing implications of large flows within the intermediary sector. I present a general equilibrium model in an economy populated by three heterogeneous agents with recursive preferences who face state-dependent borrowing constraints.

The rest of the paper is organized as follows. Section 2 presents motivating evidence for the heterogeneity in the financial intermediary sector. Section 3 provides the theoretical framework for the general equilibrium heterogeneous-intermediary model. Sections 4 and 5 present model solution and parameter values used for calibrating the model. Section 6 provides model results. In Section 7, I study the empirical implications of the model for time-series return predictability and the cross-section of asset returns. Section 8 concludes.

2 Motivating Evidence

Before presenting the theoretical framework, in this section, I provide motivating evidence on heterogeneity of the intermediary sector. Empirical evidence from asset reallocations within the financial sector recorded during the Great Recession seem puzzling through the lens of representative intermediary-based models.[11] Depending on the type of frictions considered, models with a representative financial sector can only describe intermediaries who are either buyers or sellers of

[11]Evidence of intermediary heterogeneity during the crisis has also been recently documented in the literature. He, Khang, and Krishnamurthy (2010) and Begenau et al. (2019) document flows of financial assets within the intermediary sector during the Great Recession and show that broker-dealers and hedge funds reduced leverage by selling securitized assets to commercial banks who have access to stable deposits. Ang, Gorovyy, and van Inwegen (2011) document that hedge funds decreased leverage prior to the onset of the financial crisis while the leverage of banks and the financial sector continued to increase. Ben-David, Franzoni, and Moussawi (2012) provide additional evidence of hedge fund deleveraging during the crisis.

assets during a crisis, but not both, implying opposite cyclical dynamics for intermediary leverage: *countercyclical* in models with equity constraints (He and Krishnamurthy, 2012, 2013 and Brunnermeier and Sannikov, 2014) or *procyclical* if intermediary faces a debt constraint (Brunnermeier and Pedersen, 2009 and Adrian and Shin, 2014).

Intermediaries exhibit heterogeneous behavior in the cyclical properties of their leverage. Figure 1a presents time-series of leverage for different financial intermediaries: security brokers and dealers (BDs), and bank holding companies of New York Fed's primary dealers (BHCs), intermediaries recently studied in AEM, and HKM, respectively.[12] Broker-dealer's (book) leverage is calculated from balance sheet data in Table L.130 of the Financial Accounts of the United States (Flow of Funds) from the Federal Reserve and is defined as the ratio total financial assets to total equity (total financial assets minus total liabilities). BHC leverage is defined as the ratio of total market assets (book debt plus market equity) to total market equity constructed for publicly-traded holding companies of the New York Fed's primary dealer counterparties using data from CRSP/Compustat and Datastream. Over the period from the first quarter of 2008 to the fourth quarter of 2009, which includes the Lehman bankruptcy in the fall of 2008, broker-dealers *reduced* leverage by approximately 47% (from 35 to 19) while holding companies *increased* leverage by approximately 72% (from 22 to 38) during the same period.[13] We observe opposite cyclical leverage patterns for different financial intermediaries: BD leverage is *procyclical*, while BHC leverage is *countercyclical*. Over the sample period from 1970Q1 to 2017Q4, shocks to broker-dealer leverage exhibit a *positive* correlation of 0.12 (t-stat of 1.82) to innovations in the real GDP, while BHC leverage shocks have a *negative* correlation of -0.19 (t-stat of -2.62). Correlations with GDP innovations become stronger post 2000 with coefficients of 0.37 (t-stat of 3.27) and -0.39 (t-stat of -3.56) for broker-dealer and holding company leverage, respectively.[14]

[12] A bank holding company (BHC) is a corporation which controls one or more banks. A typical U.S. prents BHC owns a number of deposit-taking bank subsidiaries and also other non-banking and foreign subsidiaries engaged in securities dealing, underwriting, insurance, real estate, etc. For example, Citibank is a commercial bank owned by Citigroup, which is its parent BHC that also owns a broker-dealer subsidiary (Citigroup Global Markets Inc.) among other non-banking and foreign subsidiaries. See Avraham, Selvaggi, and Vickery (2012) for more details.

[13] According to He et al. (2010), book leverage of commercial banks rose from 10 to between 20 and 32 over the period from 2007Q4 to 2009Q1.

[14] As discussed in HKM, observed opposite cyclical properties for leverage of different intermediaries above are unlikely to be entirely attributed to the differences between book- and market-based values for calculating BD and BHC leverage, respectively. To see this, I calculate holding company *book* leverage by simply replacing market equity with book equity in the calculation above. I find that book and market BHC leverage are in fact strongly *positively*

In Figure 1b, I plot the quarterly change in total financial assets for security broker-dealers and private depository institutions (commercial banks) from Tables L.130 and L.110 of the Flow of Funds, respectively. Over the period from the first quarter of 2008 to the fourth quarter of 2009, broker-dealers, who mainly depend on collateralized repo financing, massively *reduced* asset holdings by approximately \$1.7 trillion (from \$4.9 to \$3.2 trillion), while commercial banks, who have access to more stable deposit financing, *increased* total asset by nearly \$1 trillion (from \$13.4 to \$14.4 trillion). This evidence is consistent with findings of He et al. (2010) who document that during the 2008 crisis, hedge funds and broker-dealers reduce holdings of securitized assets by approximately \$800 billion and commercial banks increase holdings of these assets by approximately \$550 billion.[15]

Models with representative intermediaries are unable to capture this heterogeneity within the financial sector and study its implications for asset prices and the real economy. In the next section, I present a general equilibrium model with heterogeneous intermediaries and financial frictions that is consistent with opposite cyclical dynamics of leverage within the financial sector. My model implies that a dealer deleveraging episode comparable to one observed during the recent financial crisis, leads to an approximately 55% increase in the risk premia and a 5% increase in endogenous volatility. I then study model's asset pricing implications for time-series predictability and the cross-section of expected returns.

3 Model

In this section I present a general equilibrium model featuring heterogeneous intermediaries and financial constraints. My model reconciles seemingly contradictory results for the sign of price of intermediary leverage shocks from recent empirical evaluations of representative intermediary-based models. The model nests key forces behind Brunnermeier and Pedersen (2009) and He and Krishnamurthy (2012) with two main ingredients: (i) agents differ in their attitudes toward risk,

correlated over the sample period (1970Q1 to 2017Q4) with correlation coefficient of 0.64 (t-stat of 11.64). Also, both market and book leverage for BHCs are strongly *negatively* correlated with book leverage of broker-dealers over the sample period with correlation coefficients of -0.50 (t-stat of -7.89) and -0.24 (t-stat of -3.47), respectively.

[15] From the fourth quarter of 2007 to the first quarter of 2009, the Federal Reserve and the GSEs increased holdings of securitized assets by approximately \$350 billion. See He et al. (2010) for more details.

and (ii) different intermediaries face state-dependent leverage constraints that occasionally bind, while equity issuance is ruled out by assumption.[16]

To study the implications of heterogeneity in the financial sector, I augment a heterogeneous-agent asset pricing model to develop a parsimonious dynamic framework that represents the differences between these two intermediaries in a simple way. In the model, different intermediaries are represented by agents, heterogeneous in their risk-bearing capacity, who run these institutions. Ideally, one would prefer a richer framework which features the institutional details of these intermediaries. In my model, I abstract away from many of these institutional details for two main reasons: (i) solving such a model proves to be an extremely challenging task, and more importantly, (ii) the publicly available aggregate intermediary balance sheet data that I use do not help shed light on the asset pricing implications of these institutional features. I believe addressing the implications of these features in a richer framework with detailed institutional data sets is a fruitful area for future research.

I consider an endowment economy in continuous time populated by a continuum of agents whose total mass is one.[17] There are three types of agents: A, B, and C with recursive preferences and different levels of risk aversion who face state-dependent margin constraints. While I extend a heterogeneous-agent model with occasionally binding leverage constraints to a setting with three agents (with two levered intermediaries) and recursive preferences, the main contribution of my paper is: (i) presenting a dynamic framework to quantitatively analyze the importance of heterogeneity in the financial sector, (ii) showing it is able to match patterns of heterogeneity observed in the data, (ii) finally empirically validating its asset pricing predictions.

To ensure the existence of a non-degenerate stationary wealth distribution, I assume each agent faces an exogenous constant mortality rate $\kappa > 0$. New agents are born at the same rate κ per unit of time with a fraction $\bar{u}$ as type A, a fraction $\bar{v}$ as type B, and a fraction $1-\bar{u}-\bar{v}$ as type C. So, the total population is kept constant (normalized to one).[18] In aggregate, the newborns inherit the

[16]As mentioned above, in the baseline calibration of the model, leverage constraint only occasionally bind for the most risk-tolerant agent.

[17]The model can be easily extended to an AK production economy which allows for capital accumulation, investment, and real effects.

[18]By the law of large numbers, at time t a fraction $\kappa e^{-\kappa(t-s)}$ of agents born at time $s \leq t$ survive and thus the total population at time t is equal to $\int_{-\infty}^{t} \kappa e^{-\kappa(t-s)} ds = 1$.

wealth of their deceased parents on a per capita basis. Gârleanu and Panageas (2015) show that under these conditions, the possibility of exit makes agents more impatient: their effective time preference is increased by κ (i.e., from ρ to $\rho + \kappa$).[19]

3.1 Endowment and Agents

The aggregate endowment D_t evolves according to

$$\frac{dD_t}{D_t} = \mu_D \, dt + \sigma_D \, dZ_t, \tag{1}$$

where μ_D and σ_D are constant parameters and Z_t is a standard Brownian motion defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\{\mathcal{F}_t, t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ satisfying the usual conditions, as defined by Protter (2004).[20] The shock dZ_t is the only source of uncertainty in the model representing a permanent shock to the aggregate dividend. I assume that the growth rate of the endowment is positive, $\mu_D - \sigma_D^2/2 > 0$. Without loss of generality I set $D_0 = 1$. Similar to He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014), I assume agents are unable to hedge the aggregate risk.[21]

To separate the effects of elasticity of intertemporal substitution (EIS) and risk aversion, I assume all agents have stochastic differential utility as in Duffie and Epstein (1992), the continuous-time analog of recursive preferences of Epstein and Zin (1989). In particular, an agent of type i has the lifetime utility $U_{i,t}$ at time t given by

$$U_{i,t} = \mathbb{E}_t \left[\int_t^\infty f_i \left(C_{i,s} \, , U_{i,s} \right) ds \right], \tag{2}$$

where

$$f_i(C_{i,t}, U_{i,t}) = \left(\frac{1 - \gamma_i}{1 - 1/\psi_i} \right) U_{i,t} \left[\left(\frac{C_{i,t}}{[(1 - \gamma_i) U_{i,t}]^{1/(1-\gamma_i)}} \right)^{1 - 1/\psi_i} - (\rho + \kappa) \right]. \tag{3}$$

Function f_i aggregates over current consumption $C_{i,t}$ and future utility $U_{i,t}$. Parameters γ_i and ψ_i

[19] The only purpose of introducing the OLG framework and mortality risk is to make the model stationary.

[20] The filtration represents the resolution over time of information commonly available to investors.

[21] Di Tella (2017) provides a moral hazard framework where aggregate uncertainty shocks lead to balance sheet recessions even though agents can write complete contracts on the aggregate state of the economy.

denote agent i's coefficient of relative risk aversion and EIS, respectively. These preferences reduce to standard power utility when $\psi_i = 1/\gamma_i$. All agents are assumed to have a common subjective discount factor ρ increased by κ as mentioned above.

Agents are heterogeneous in their attitudes toward risk γ_i. A agents are the most risk tolerant and C agents are the most risk averse. B agents are more risk tolerant than C types: $\gamma_A < \gamma_B < \gamma_C$. I think of A agents representing shadow banks (broker-dealers, hedge funds, etc.) and B agents as traditional banks, and C agents representing the household sector. In equilibrium A and B will have levered balance sheets by borrowing from C agents. The financial sector (A and B agents) face time-varying margin constraints, which I discuss later in detail.

3.2 Financial Markets and Budget Constraints

All agents can trade a risky asset in fixed supply (normalized to one) and an instantaneous (from t to $t + dt$) risk-free bond in zero net supply which pays the endogenously-determined interest rate r_t. The risky asset is a claim on the aggregate endowment $\{D_t\}$, so, the total return on the risky claim is

$$dR_t = \frac{dP_t + D_t\,dt}{P_t} \equiv \mu_t\,dt + \sigma_t\,dZ_t, \tag{4}$$

where P_t is the price of the risky claim, μ_t is its expected return, and σ_t is its volatility, all determined in equilibrium. I use the consumption good as the numeraire. I also denote the dividend-price ratio of the risky asset by $F_t = D_t/P_t$.

Let $W_{i,t}$ denote agent i's wealth and assume $W_{i,0} > 0$ for $i \in \{A, B, C\}$.[22] Let w_s^i be the share of agent i's wealth invested in the endowment claim. Then agent i's financial wealth evolves according to the following standard dynamic budget constraint

$$\frac{dW_{i,t}}{W_{i,t}} = \left(r_t + w_{s,t}^i\left(\mu_t - r_t\right) - c_{i,t}\right)dt + w_{s,t}^i \sigma_t dZ_t, \tag{5}$$

where $c_i \equiv C_i/W_i$ is agent i's consumption-wealth ratio. The agent earns the risk-free rate, earns the risk premium on the risky asset, and pays for consumption. The intermediary leverage is

[22] Throughout the paper, I use terms net worth and equity interchangeably.

defined as the ratio of asset over equity. Thus, when portfolios weights w_s^A or w_s^B exceeds one, the intermediaries operate with leverage by raising debt from households C.

3.3 Financial Constraints

I assume agent face a occasionally binding state-dependent margin constraint: at each moment in time, borrowers are restricted on how much leverage they can use on their balance sheets. In other words, lenders impose margin requirements to protect themselves against losses caused by adverse price movements.[23]Margins are set to shield lenders against adverse price movements and are widely used in the financial sector to fund levered balance sheets. They have also been previously studied in the asset pricing literature (see Brunnermeier and Pedersen, 2009, Gârleanu and Pedersen, 2011, Chabakauri, 2013, and Rytchkov, 2014 for some recent examples).

The tightness of the margin constraint can be determined by the regulators (e.g. Federal Reserve Regulation T) or by security broker-dealers (e.g. overcollaterlization of repos by a hedge fund's prime brokerage).

At any time t, I assume margin constraints restricts agent i's portfolio weight w_s^i to be below a certain state-dependent threshold $\bar{\theta}_t$

$$w_{s,t}^i \leq \bar{\theta}_t, \tag{6}$$

where, $\bar{\theta}_t$ determines the form of margin constraints, which is linked to endogenously-determined equilibrium objects (e.g. volatility of risk asset returns). Since equilibrium objects also depend on the state of the economy, margin requirements are state-dependent as well.

In particular, I assume margin requirements depend on the volatility of the risky asset return σ_t, and have the following functional form

$$\bar{\theta}_t = \bar{m} \left(\frac{1}{\bar{m} \alpha \sigma_t} \right)^{\nu}, \tag{7}$$

where ν, α, and $\bar{m}$ are parameters that determine the type and tightness of the constraint, respec-

[23]In this paper I abstract from the question of why the intermediaries face dynamic margin constraint and do not model the contracting problem among agents. Adrian and Shin (2014) provide a microfoundation for the Value-at-Risk constraint using a moral hazard problem in a static partial equilibrium setting.

tively. When $\nu = 0$, agents face a *constant* margin requirement: $\bar{\theta} = \bar{m}$. When $\nu = 1$, equation (7) resembles a Value-at-Risk rule.[24] In the latter case, the level of margin constraint is endogenous since it is inversely linked to the return volatility, which is an equilibrium object.[25]

As mentioned earlier and consistent with empirical evidence, although all agents face margin constraints, in equilibrium, only more aggressive A type intermediaries face constraints that occasionally bind.[26]

3.4 Agents' Optimization Problems

Since agents are identical within each type and have homothetic preferences, I consider the problem faced by a representative agent i for $i \in \{A, B, C\}$. Each agent solves a standard Merton (1973) dynamic portfolio choice problem subject to margin constraints: agent i starts with initial wealth $W_{i,0} > 0$, decides how much to consume as a fraction of her wealth, $c_{i,t}$, and what fraction of her net worth to invest in risky asset, $w^i_{s,t}$, in order to maximize her value function in (2), subject to the dynamic budget constraint (5) and endogenous margin constraints (6). So, agent i's problem is

$$V_{i,t} = \max_{(c_i \geq 0, w^i_s)} U_{i,t}$$

$$\text{s.t.:} \quad \text{dynamic budget constraint (5) and margin constraint (6)} \tag{8}$$

and a solvency constraint $W_{i,t} \geq 0$.

3.5 Equilibrium

The definition of the competitive equilibrium is standard and is given below.

Definition 1. *A competitive equilibrium is the set of aggregate stochastic processes adapted to the filtration generated by Z_t: the price of claim on the aggregate endowment P, and the risk-free*

[24] Value-at-Risk constraints aim at limiting downside risk and maintaining an equity cushion large enough so that the default probability is kept below some benchmark level. They are common for banks and other leveraged financial institutions and are embedded in Basel II and Basel III regulatory frameworks. See Danielsson et al. (2012) and Adrian and Shin (2014) for recent examples.

[25] In equilibrium, more risk averse C agents lend to levered intermediaries. Thus, given the my calibration, the constraint can potentially bind only for A and B types.

[26] This assumption can be relaxed by allowing the margin constraint to bind for B agents (banks) as well without affecting the main results of the model.

interest rate r; and a set of stochastic processes for each agent i: net worth W_i, consumption C_i, and stock holdings w_s^i; such that:

i. Given the aggregate stochastic processes (P_t, r_t), choices $\left(C_{i,t}, w_{s,t}^i\right)$ solve agent i's optimization problem in (8).

ii. Markets clear

$$C_{A,t} + C_{B,t} + C_{C,t} = D_t \quad \textit{(goods market)} \tag{9}$$

$$w_{s,t}^A W_{A,t} + w_{s,t}^B W_{B,t} + w_{s,t}^C W_{C,t} = P_t \quad \textit{(stock market)} \tag{10}$$

The bond market clears by Walras' law. Note that bond market clearing implies that the aggregate wealth in the economy is equal to the value of the endowment claim, i.e.

$$W_{A,t} + W_{B,t} + W_{C,t} = P_t \, .$$

4 Model Solution

In order to solve the model, I need to determine how prices, portfolio choices, and consumption processes for all agents depend on the historical paths of the aggregate shock Z_t. The equilibrium can be characterized in a recursive formulation where all equilibrium objects are functions of two endogenous state variables, defined below. The computation of equilibrium requires solving the Hamilton-Jacobi-Bellman (HJB) partial differential equations of A, B, and C agents simultaneously. Unfortunately, the system of nonlinear PDEs does not admit a closed-form solution and I have to rely on numerical techniques. In this section, I first define my model's two endogenous state variables and derive their dynamics. I then characterize agents' value functions and provide some intuition for their optimal portfolio and consumption policy functions. I define a recursive Markov equilibrium and finally briefly discuss the numerical algorithm used to solve the PDEs.

4.1 Endogenous State Variables

Because Epstein-Zin preferences are homothetic, the optimal control variables for an agent are all linear in her wealth. The linear property allows me to simplify the endogenous state space, from an infinite-dimensional into a two-dimensional space. More precisely, I only need to keep track of the share of aggregate wealth that belongs to types A and B (the financial sector), as well as, the wealth share of A agents in the financial sector. I can derive equilibrium conditions as functions of the following endogenous state variables:

$$x_t \equiv \frac{W_{A,t} + W_{B,t}}{P_t}, \qquad y_t \equiv \frac{W_{A,t}}{W_{A,t} + W_{B,t}}. \tag{11}$$

Since the risk-free asset is in zero net supply, the aggregate wealth in the economy is equal to the risky asset price P_t. The state variable x is the share of aggregate wealth that belongs to the financial sector (i.e. A and B agents), and y is the type A intermediaries' wealth share as a fraction of the total financial sector.[27]

The state variable x (total wealth share of the financial sector in the economy), is the key state variable in recent intermediary asset pricing models with a representative financial sector (see He and Krishnamurthy, 2013, Brunnermeier and Sannikov, 2014, and Gertler and Kiyotaki, 2010, for example) If only intermediaries can invest in the risky asset, state variable x represents the equity capital ratio of the financial sector.[28] HKM show shocks to capital ratio of intermediaries price the cross-section of expected return with a positive price of risk: intermediary's marginal value of wealth rises when capital ratio x falls.

State variable y, on the other hand, captures the wealth distribution within the intermediary sector. It represents heterogeneity among intermediaries in the sense that it would not be present in models with a representative financial sector. Distribution of wealth among different intermediaries clearly plays no role in the models with a representative financial sector. In contrast, in Section 7, I show that the distribution of wealth between broker-dealers and bank holding companies (proxies for A and B agents, respectively) can negatively forecasts future returns for many asset classes. I

[27] Note that the definitions in equation (11) ensure that the domain of both state variables is $[0, 1]$.

[28] In this case, because riskless bonds are in zero net supply and the risky asset is assumed to be in unit supply, total assets of the intermediary sector is equal to the risky asset price P.

also demonstrate that shocks to y are a priced risk factor in the cross-section of equity and bond returns with a positive estimated price of risk.

The representation above is not the only way one can construct the state variables in the model. My choice of state variables x and y however, presents a natural way to do so: in many representative intermediary-based models, x is the main state variable. Moreover, with the choice of y presented here, in the absence of heterogeneity, my model collapses to one with x as the only state variable similar to the models with a representative financial sector. So this presents a natural way to construct the state variables in the model.

I restrict my attention to a Markov equilibrium (defined below) in the state space $(x, y) \in [0,1] \times [0,1]$, where all processes are functions of (x_t, y_t) only. Proposition 1, characterizes the dynamics of the two endogenous state variables (x, y).

Proposition 1. *The laws of motion for endogenous state variables x and y are given by*

$$dx_t = \kappa\left(\bar{x} - x_t\right) dt + x_t \left(\mu_{x,t}\, dt + \sigma_{x,t}\, dZ_t\right), \tag{12}$$

$$dy_t = \kappa\left(\bar{y} - y_t\right) dt + y_t(1 - y_t)\left(\mu_{y,t}\, dt + \sigma_{y,t}\, dZ_t\right) \tag{13}$$

where $\bar{x} = \bar{u} + \bar{v}$ and $\bar{y} = \bar{u}/(\bar{u} + \bar{v})$.

i. The drifts of x and y are given by

$$\mu_x = \left[y w_s^A + (1-y) w_s^B - 1\right]\left(\mu - r - \sigma^2\right) - y c_A - (1-y) c_B + F \tag{14}$$

$$\mu_y = \left(w_s^A - w_s^B\right)(\mu - r) - c_A + c_B - \left[y w_s^A + (1-y) w_s^B\right]\left(w_s^A - w_s^B\right)\sigma^2 \tag{15}$$

ii. The diffusions of x and y are given by

$$\sigma_x = \left[y w_s^A + (1-y) w_s^B - 1\right]\sigma \tag{16}$$

$$\sigma_y = \left(w_s^A - w_s^B\right)\sigma \tag{17}$$

Proof. See Appendix A. □

Given the dividend-price ratio F, the return process for the endowment claim in equation (4) can be rewritten as

$$dR = \frac{d(D/F)}{D/F} + F\,dt = \mu\,dt + \sigma\,dZ,$$

where time subscripts are dropped for notational simplification.

Using Ito's lemma, the expected return and volatility of the risky asset will be

$$\begin{aligned}
\mu &= \mu_D + F - \frac{F_x}{F}\left[\kappa(\bar{x} - x) + x(\mu_x + \sigma_D\sigma_x)\right] - \frac{F_y}{F}\left[\kappa(\bar{y} - y) + y(1-y)(\mu_y + \sigma_D\sigma_y)\right] \\
&\quad + \left[\left(\frac{F_x}{F}\right)^2 - \frac{1}{2}\frac{F_{xx}}{F}\right] x^2\sigma_x^2 + \left[\left(\frac{F_y}{F}\right)^2 - \frac{1}{2}\frac{F_{yy}}{F}\right] y^2(1-y)^2\sigma_y^2 \\
&\quad + \left[2\left(\frac{F_x}{F}\right)\left(\frac{F_y}{F}\right) - \frac{F_{xy}}{F}\right] xy(1-y)\sigma_x\sigma_y \qquad (18) \\
\sigma &= \sigma_D - \frac{F_x}{F}x\sigma_x - \frac{F_y}{F}y(1-y)\sigma_y \qquad (19)
\end{aligned}$$

Note that from (19), a part of the risk from holding the risky asset is fundamental, $\sigma_D\,dZ_t$, and a part is *endogenous*, $\left(-\frac{F_x}{F}x\,\sigma_x - \frac{F_y}{F}y(1-y)\,\sigma_y\right)dZ_t$. Equation (19) also implies that the volatility of returns σ exceeds the fundamental volatility σ_D when price-dividend ratio, $1/F$, and the state variables x and y are procyclical, i.e. $F_x > 0, F_y > 0, \sigma_x > 0$, and $\sigma_y > 0$, which is the case in equilibrium.

The following proposition provides the boundary conditions that the state variable diffusions satisfy.

Proposition 2. *The diffusion for state variables* (x_t, y_t) *satisfy the following boundary conditions:*

$$\lim_{x\to 0} x\,\sigma_{x,t} = \lim_{x\to 1} x\,\sigma_{x,t} = 0, \quad \forall y \in [0,1]$$
$$\lim_{y\to 0} y(1-y)\,\sigma_{y,t} = \lim_{y\to 1} y(1-y)\,\sigma_{y,t} = 0, \quad \forall x \in [0,1]$$

Proof. See Appendix A. □

These boundary conditions will be used later to solve agents' HJB equations discussed below.

4.2 Hamilton-Jacobi-Bellman Equations

The recursive formulation of agent i's optimization problem is given by the following HJB equation

$$0 = \max_{c_i, w_s^i} f_i(c_i W_i, V_i(W_i, x, y))dt + \mathbb{E}_t \left[dV_i(W_i, x, y) \right], \tag{20}$$

where V_i is agent i's value function. With homothetic preferences, the value functions have the power form. The Following proposition characterizes agents' value functions.

Proposition 3. *The value function of agent $i \in \{A, B, C\}$ has the form*

$$V_i(W, x, y) = \frac{W_i^{1-\gamma_i}}{1-\gamma_i} J_i(x, y)^{\frac{1-\gamma_i}{1-\psi_i}}, \tag{21}$$

where J_i is agent i's consumption-wealth ratio, $c_i = J_i$.

Furthermore, J_i solves the following second-order partial differential equation (PDE)

$$\begin{aligned}
\rho + \kappa = & \frac{1}{\psi_i} J_i + \left(1 - \frac{1}{\psi_i}\right) \left[r + w_s^i(\mu - r) - \frac{\gamma_i}{2} \left(w_s^i\right)^2 \sigma^2 \right] \\
& - \frac{1}{\psi_i} \left\{ \frac{J_{i,x}}{J_i} \left[\kappa(\bar{x} - x) + x\mu_x \right] + \frac{J_{i,y}}{J_i} \left[\kappa(\bar{y} - y) + y(1-y)\mu_y \right] \right. \\
& \left. + (1-\gamma_i) \left(\frac{J_{i,x}}{J_i} x\sigma_x + \frac{J_{i,y}}{J_i} y(1-y)\sigma_y \right) w_s^i \sigma \right\} \\
& - \frac{1}{2\psi_i} \left[\left(\frac{\psi_i - \gamma_i}{1 - \psi_i} \right) \left(\frac{J_{i,x}}{J_i} x\sigma_x + \frac{J_{i,y}}{J_i} y(1-y)\mu_y \right)^2 + \frac{J_{i,xx}}{J_i} x^2 \sigma_x^2 \right. \\
& \left. + 2 \frac{J_{i,xy}}{J_i} xy(1-y)\sigma_x \sigma_y + \frac{J_{i,yy}}{J_i} y^2 (1-y)^2 \sigma_y^2 \right]
\end{aligned} \tag{22}$$

Proof. See Appendix A. □

Functions J_i capture agent i's investment opportunity set. In particular, note that from (21) if $\frac{1-\gamma_i}{1-\psi_i} > 0$ (which holds in my calibration), marginal utility of wealth is increasing in J_i.

The first-order conditions of agent's recursive problem gives the optimal consumption and portfolio choice

$$c_i = \frac{C_i}{W_i} = J_i \tag{23}$$

$$w_s^{i,*} = \frac{\mu - r}{\gamma_i \, \sigma^2} + \frac{1}{\gamma_i}\left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i} x \, \frac{\sigma_x}{\sigma} + \frac{J_{i,y}}{J_i} y(1-y)\frac{\sigma_y}{\sigma}\right) \tag{24}$$

The optimal unconstrained portfolio $w_s^{i,*}$ is the standard ICAPM result of Merton (1973): the first term in (24) is the myopic demand of a one-period mean-variance investor and the second term is the hedging demand capturing the variations in the agent's investment opportunity set. The optimal consumption-wealth ratio in (23) comes from the standard envelope condition.

So, from the optimal portfolio in the absence of constraints in (24) and the margin constraint in equation (6), the optimal portfolio is

$$w_{s,t}^i = \min\left(\bar{\theta}_t, w_{s,t}^{i,*}\right), \tag{25}$$

where the leverage upper bound $\bar{\theta}_t$ is defined in (7).

4.3 Recursive Markov Equilibrium

I derive a Markov equilibrium in state variables x_t and y_t. That is, I look for an equilibrium where all equilibrium objects (prices, consumption, and portfolio choices) can be written as functions of these two state variables. Next I define the Markov equilibrium in state space (x, y).

Definition 2. *A Markov equilibrium in state variables (x_t, y_t) is the set of functions: marginal value of wealth $J_i(x, y)$, dividend-price ratio $F(x, y)$, real interest rate $r(x, y)$ and policy functions $c_i(x, y), w_s^i(x, y)$ for $i \in \{A, B, C\}$, and laws of motion for endogenous state variables $\mu_x(x, y), \mu_y(x, y)$ and $\sigma_x(x, y), \sigma_y(x, y)$ such that*

i. marginal value of wealth J_i solves agent i's HJB equation, and c_i and w_s^i are corresponding policy functions, taking F, r and laws of motion for x and y as given.

ii. Markets for consumption good and risky asset clear

$$xy\, c_A + x(1-y)\, c_B + (1-x)\, c_C = F \quad (\textit{goods market}) \tag{26}$$

$$xy\, w_s^A + x(1-y)\, w_s^B + (1-x)\, w_s^C = 1 \quad (\textit{stock market}) \tag{27}$$

iii. The laws of motion for x and y satisfy (14)–(17).

4.4 Numerical Solution

The model is analytically tractable: the equilibrium dynamics can be fully characterized by a system of partial differential equations that are solved numerically. The computation of equilibrium requires solving the HJB equations of the three types of agents simultaneously. Functions $J_A(x,y)$, $J_B(x,y)$, and $J_C(x,y)$ can be found by solving a system of second-order partial differential equations (PDEs) in (x,y). To do so, all equilibrium objects (e.g. $F, \sigma, \mu, \sigma_x, \mu_x, \sigma_y, \mu_y$, etc.) need to be expressed in terms of functions J_i and their derivative. Unfortunately, the system of nonlinear differential equations does not admit a closed-form solution and I have to rely on numerical techniques. This is particularly challenging in the presence of model's two endogenous state variables. I use projection methods, specifically orthogonal collocation using Chebyshev polynomials (Judd, 1992, 1998), to solve for equilibrium. Unlike a log-linearized representation around the steady state, this method provides a global solution and a full characterization of the whole dynamic system. In Appendix B, I explain the numerical procedure in detail.

5 Calibration and Parameters Choices

Table 1 lists the parameter choices used in calibrating the model. While my main goal is to illustrate the mechanisms of the model, I pick parameter values that I view as reasonable. Note that since my model is set in continuous time, the parameter values in Table 1 correspond to annual values rather than the typical quarterly values used in calibrating discrete-time macro models. I choose the drift and diffusion of the aggregate endowment (μ_D and σ_D) so that time-integrated data from the model can roughly match the first two moments of annual U.S. consumption growth. I set $\mu_D = 0.022$ and $\sigma_D = 0.035$ consistent with the long historical sample of Campbell and Cochrane (1999). Parameter κ, which controls the entry and exit of agents, is set to 0.0154. This number is close to the US birth rate from 1970 to 2015.[29] The value of κ implies that agents on average live for 65 years after they start making economic choices. Assuming this age is about 20, $\kappa = 0.0154$ implies an average lifespan of 85 years, consistent with the calibration in Gârleanu and Panageas (2015). The subjective discount rate ρ is set to 0.001 which results in real interest rates of between 2.5% and 3% annually. Note that with this value of ρ, the investors' effective discount rate is $\rho + \kappa = 0.016$, comparable to calibrations used in the asset pricing literature.[30]

The parameter $\bar{m}$ in equation (7) determines the constant margin requirements (when $\nu = 0$), and I set $\bar{m} = 4$, the same order of magnitude as Rytchkov (2014). When $\nu = 1$, equation (7) resembles a Value-at-Risk constraint. In this case, parameter α determines the tightness of the constraint. I set $\alpha = 10$ which is approximately equal to the one-month Value-at-Risk at the 99% level.

I set a common value of EIS for all agent types of $\psi_i = \psi = 1.5$, similar to the values in Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012) and others. I treat the risk aversion of A intermediaries as a free parameter and use the remaining parameters to approximately match leverage and asset pricing moments. To match the ratio of broker-dealer leverage (A intermediaries) to that of banks (B intermediaries), unconditional Sharpe ratio, and create substantial demand for risk sharing and leverage, I set the risk aversions of A, B, and C agents to 2.5, 5.5, and 15, respectively. From the Flow of Funds, in my sample period (1970Q1-2017Q4), the average leverage of broker-dealers is approximately 2.2 times higher than that of banks, roughly what I get in the model when state variables are at their unconditional means. The choice of risk aversion coefficients gives an average Sharpe ratio of the model of 38%, which is in the range of typical consumption-based asset pricing calibrations.

The values of γ_i and ψ_i imply agent i's preference for the early resolution of uncertainty and have been extensively used in the asset pricing literature to address a number of asset pricing puzzles.[31] A value of EIS greater than one implies a decline in asset prices when the effective risk aversion in the economy increases. The shares of type A and B agents in the population, $\bar{u}$ and

[29]Source: National Vital Statistics Report, Volume 66, Number 1, January 5, 2017.

[30]For instance, agents' effective discount rate in this paper is close to the calibrations in Gârleanu and Panageas (2015) and Drechsler et al. (2018).

[31]See, for example, Bansal and Yaron (2004), Hansen, Heaton, and Li (2008), and Bansal and Shaliastovich (2013) for resolution of equity premium, value premium, and uncovered interest rate parity puzzles, respectively.

$\bar{v}$, are chosen to target share of intermediaries net worth as well as dealer's net worth share in the financial sector.

6 Model Results

In this section, I first present additional properties of the equilibrium with margin constraints and compare them with the unconstrained economy with the same fundamentals and degree of heterogeneity among agents. The economy with margin constraints simultaneously exhibits higher risk premium and lower risk-free rate and volatility, compared to the frictionless benchmark. Although some of these effects have been previously documented in the literature, my analysis extends these results to an economy with three agents (households and two heterogeneous intermediaries) and recursive preferences.[32] The equilibrium with three heterogeneous agents and two endogenous state variables is considerably more challenging to solve numerically.

In Section 6.2, I show, consistent with empirical evidence, the model can generate different cyclical dynamics for different intermediaries (i.e. A and B agent). Implications of heterogeneity in the financial sector (captured by the wealth distribution among intermediaries) for variation is risk premia are discussed in Section 6.3. Section 6.4 studies the impact of a dealer deleveraging episode comparable to the one observed during the recent financial crisis, for risk premia and volatility. Finally, in Section 6.6, I show how the model reconciles seemingly contradictory evidence for the sign of price of intermediary leverage shocks in AEM and HKM.

6.1 Constrained versus Unconstrained Economy

Unconstrained Benchmark

As a benchmark, I first consider an economy without margin constraints, that is $\bar{\theta}(x_t, y_t) \to \infty$. In the absence of constraints, investors face complete markets and their Euler equations hold with equality in equilibrium: an economy is very similar to Gârleanu and Panageas (2015) but with *three* heterogeneous agents.

[32]For example, Rytchkov (2014) adds margin constraints to an endowment economy with two heterogeneous agents and CRRA preferences (similar to the models in Longstaff and Wang, 2012 and Bhamra and Uppal, 2014) to show binding constraints reduce return volatility and risk-free rate, but increase expected returns and Sharpe ratio.

Figures 2 presents various equilibrium variables (price-dividend ratio $1/F$, volatility of the risky asset return σ, Sharpe ratio $(\mu - r)/\sigma$, and risk premium on the endowment claim $\mu - r$) as a function of the state variable x while the state variable y is fixed at (its stochastic steady state). Three-dimensional plots for equilibrium objects are provided in Appendix C.2. In both figures solid blue lines correspond to the frictionless economy. Along the horizontal axis in each panel is the state variable x_t (the wealth share of types A and B), which ranges from 0 to 1.

The top right panel of Figure 2 shows the volatility of returns. Even though fundamental volatility is constant ($\sigma_D = 3.5\%$ in my calibration), return volatility is time varying and it *exceeds* fundamental volatility in a hump-shaped pattern. As mentioned above and shown in equation (19), a part of the risk from holding the risky asset is fundamental and a part is endogenous. From equation (11), wealth shares of agents A, B, and C in the aggregate economy are equal to $x\,y, x(1-y)$ and $1 - x$, respectively. Thus, when $(x, y) = (1, 1)$, $(x, y) = (1, 0)$, or $(x = 0, \forall y)$, the economy is populated by one type of agent (A, B, and C, respectively) and the volatility of the endowment claim coincides with the fundamental volatility σ_D.[33] The Sharpe ratio and expected excess return $\mu - r$ both show countercyclical behavior as expected: higher risk premium and price of risk during distressed states when the intermediary sector is undercapitalized (low-x states) and/or broker-dealers deleverage (when y is low) are low.

The bottom right panel of Figure 2 shows that the risk premium largely tracks the Sharpe ratio. Note that this is the risk premium on a claim to the aggregate endowment, which has a relatively low volatility (3.5% in the baseline calibration). By comparison, equity volatility is around 16%. Therefore, the equity premium implied by the model is about five to six times larger than that of the endowment claim, putting it in the range of standard estimates in the literature.

When the EIS exceeds one, the substitution effect dominates the income effect so that greater risk aversion reduces asset demand and valuations fall. In this case the rise in the risk premium exceeds the fall in the real rate. In contrast, if the EIS is less than one, greater risk aversion counter-intuitively causes the valuations of risky assets to increase.

Figure 3 shows optimal portfolio weights in the risky asset for all three agents. For the most risk tolerant type A investors it always exceeds one and for the most risk averse type C agents

[33] This can be validated from three-dimensional plots in Figure G.3 in the Appendix.

it is always less than one. B's optimal portfolio is greater than one for most of the state space. Thus, without constraints, financial intermediaries (type A and B agents) borrow from type C investors (households) to take a levered positions in the endowment claim. As the wealth share of the financial sector get bigger and A agents have more relative wealth in the financial sector, they borrow from traditional banks (B agents) as well.

Importantly, in the absence of constraints, the optimal leverage of A and B investors are both *countercyclical*: they are *higher* in bad states when investment opportunities are more attractive, which is counterfactual. This means when margin constraints are imposed to limit leverage, they are more likely to bind in high marginal utility states.

The relationship between portfolio weights and the wealth shares of the financial sector (x) and wealth share of broker-dealers in the financial sector (y) in Figure 3 is the result of market clearing for the risky asset in equation (27). When x is close to zero or both x and y are close to one, a single type of agent (type C in the first case and type A in the second case) dominates the economy, which reduces the opportunity for risk sharing. In the absence of constraints, agents of the dominant type must hold all their wealth in the risky asset, whereas agents of the vanishing type can be satisfied with only a small amount of borrowing or lending. Thus, when x is near zero, households (C agents) set prices and intermediaries (A and B types) take high leverage.

Equilibrium with Dynamic Margin Constraints

To study the impact of financial constraints on the equilibrium, I solve the baseline model where face margin requirements in the form given in equation (7). As noted earlier, margin constraints are occasionally binding and state-dependent. Although all agents face the margin constraints in their optimization problems, in my calibration, the wedge between risk aversions γ_A and γ_B is such that constraints occasionally bind only for the most risk tolerant A types and more risk averse agents (i.e. types B and C) do *not* face binding leverage constraints, in equilibrium.

Consistent with Kogan et al. (2007) and Rytchkov (2014), the economy with margin constraints simultaneously exhibits higher risk premium and Sharpe Ratio and lower risk-free rate and volatility, compared to the frictionless benchmark.[34]

[34] Both papers study a two-agent economy with CRRA preferences, whereas my model has three heterogeneous

I focus on two cases: (i) a constant margin constraint with $\bar{\theta}_t = \bar{m}$ ($\nu = 0$ in equation 7), and (ii) the case where type A agents face endogenous time-varying constraints ($\nu = 1$) in the form $\bar{\theta}_t = 1/(\alpha\sigma_t)$, where parameter α determines the tightness of the constraint. In the second case, margin requirements are determined by a Value-at-Risk-type rule and the level of margins is endogenous because it is (inversely) related to the return volatility, an equilibrium object.

Figures 2 and 3 present various objects for equilibria with constant (dash-dotted purple line) and state-dependent Value-at-Risk-type (dashed red line) margin constraints. There are few important observations from these figures. First, the impact of both types of constraints are qualitatively similar and the only difference is the magnitude. With the choice of parameters presented in Table 1, the time-varying margins are more restrictive and the effects are stronger with $\nu = 1$ in equation (7).

The top left panel of Figure 2 shows the impact of margin constraints on the price-dividend ratio. In my calibration, margin constraints do not substantially decrease asset's valuation ratio when they bind, reducing the price-dividend ratio by less than 1% at steady state relative to the complete-market benchmark.

The top right panel of Figure 2 plots the return volatility σ. The impact of constraints on volatility is unambiguous: portfolio constraints *reduce* the volatility of the risky asset return relative to the unconstrained economy. Also the volatility decreases in states where the constraint actually binds, although the point at which the constraint starts to bind depends on the form and severity of the constraint. The intuition behind this effect is as follows. In equilibrium, less risk averse A and B investors (the financial sector) borrow from more risk averse households and operate with leverage. It is well established in the macro-finance literature that leverage makes returns more volatile than the fundamental volatility: levered balance sheets amplify an aggregate shock to dividends.[35] Binding margin constraints reduce dynamic risk sharing and leverage in equilibrium, thereby reducing the return volatility. In my calibration, binding dynamic margin constraints results in the reduction of the return volatility by approximately 6% at model's stochastic steady state.

agents with recursive preferences and two endogenous state variables which is considerably more challenging to solve.

[35]See Kiyotaki and Moore (1997), Bernanke et al. (1999), and Brunnermeier and Sannikov (2014), for example.

The bottom panels of Figure 2 demonstrate that portfolio constraints *increase* the Sharpe ratio and risk premium of the endowment claim. This is again a general effect and does not depend on the form of the constraints. The intuition is straightforward: because the leverage of the risk tolerant agent (A type) is bounded in the part of the state space where the margin constraints bind, to clear the market, the more risk averse investors (B and C types) are forced to take on a larger portion of the risky asset that they would without constraints. To induce them to buy more, the risk premium should increase. Following negative risky asset returns, margin constraints bind and broker-dealers (A types) are forced to sell assets. As a result, to clear the market, the expected returns must increase enough to entice more risk averse agents to take on a larger supply of the risky asset than before the shock. Since banks (B types) are not facing binding constraints, as discussed above, they increase leverage following a negative shock. Thus the model can qualitatively match the empirical evidence on opposite cyclical patterns of intermediary leverage in the financial sector documented in Figure 1a. At model's stochastic steady state, binding margin time-varying constraints causes the Sharpe ratio and risk premium to rise by approximately 39% and 36%, respectively.

Figure 3 also shows the effect of margin constraints on optimal portfolio weights. As explained above, because A's leverage is countercyclical in the unconstrained economy, the margin constraints will bind in states where x and y are low. In the model with margin constraints, type A agents operate with leverage in all states of the economy, however when margin constraints bind, leverage is restricted by $\bar{\theta}_t$. In other words, the presence of binding leverage constraints makes broker-dealers' leverage *countercyclical*: they are forced to sell assets and delever in bad states of the economy where constraints bind. To clear the risky asset market, both B and C investors need to absorb this additional supply and increase their portfolio weights as we see from dashed and dotted lines in the top right and bottom left panels of Figure 3. Note that the model has difficulties generating substantial movements in and matching the level of intermediary leverage we observe during the crisis (see Figure 1a).

Finally, the middle right panel of Figure 3 also shows that margin constraints (regardless of the form) *reduce* the risk-free rate. This result is also intuitive. In the absence of constraints, A and B investors operate with leverage by borrowing from type Cs. The upper bound for leverage for type As reduces the demand for credit, thus lowering the risk-free rate. In my calibration, when margin constraint binds, the risk-free rate decreases by approximately 8% in the stochastic steady state $(x_{ss}, y_{ss}) = (0.36, 0.56)$.

Figure 4 illustrates the evolution of model's two endogenous state variables x and y in the constrained and frictionless economies. The drift of $x_t(y_t)$ is positive for low levels and becomes negative for high values of $x_t(y_t)$. The points where the drift crosses zero is the *stochastic steady state* of the endogenous state variable, the point of attraction of the system in the absence of shocks. Importantly, from the right panels in Figure 4, notice that the diffusion terms σ_x and σ_y are always positive. This implies that following a negative aggregate shock, both state variables decline, i.e. x and y both exhibit *procyclical* dynamics in the model. In Section 7.1 below, I verify this also holds in the data. As shown in Proposition 2, at the boundary points of the state space ($x = 0, x = 1, y = 0$, and $y = 1$), the diffusion of state variables x and y are zero. This can be verified from the top- and bottom-right panels of Figure 4.

The left panels of Figure 4 also illustrate that portfolio constraints of both types reduce the volatility of the state variables. Because the impact of the constraints on the sensitivity of the price-dividend ratio to the state variables is very small (as shown in the top left panel of Figure 2), a decrease in σ_x and σ_y translates into a decrease in the return volatility σ as presented in the top-right panel of Figure 2 (this follows directly from equation (19)).

6.2 Cyclical Properties of Intermediary Leverage

Countercyclical Holding Company and Financial Sector leverage, ***Procyclical*** leverage for Broker-Dealers

In this section, I show that the model is able to generate leverage patterns for different intermediaries that are consistent with the empirical evidence presented earlier. As mentioned above and illustrated in Figure 3, in the *absence* of margin constraints, broker-dealers and bank holding companies both exhibit *countercyclical* leverage: their optimal leverage is higher in bad states. However, when broker-dealers face state-dependent margin constraints inversely dependent on return volatility, their leverage exhibit an (almost) opposite cyclical behavior. Since return volatility is hump-shaped (see the top right panel of Figure 2), margin constraints cause A type's leverage to be $\cup$-shaped when constraints bind. When the constraints are sufficiently tight, shadow bank leverage is *procyclical*, consistent with the empirical evidence from broker-dealers leverage presented in Figure 1a (the solid blue line) and also documented in AEM.

Leverage of the financial sector (types A and B in the model), w_s^{FS}, is defined as the share of sector's wealth held in the risky assets:

$$w_s^{FS} = \frac{w_s^A W_A + w_s^B W_B}{W_A + W_B} = y w_s^A + (1 - y) w_s^B \tag{28}$$

where state variable $y = W_A/(W_A + W_B)$ is the wealth share of A types in the financial sector as defined in equation (11). We see that the financial sector leverage is the weighted average of A and B types' optimal leverage with a time-varying weight equal to the state variable $y \in [0, 1]$: the wealth share of broker-dealers (A types) in the financial sector.

The left panel of Figure 5 presents financial sector's leverage in the unconstrained equilibrium and in the model with endogenous margin constraints. As discussed above, margin constraints reduce financial sector's leverage when they bind: binding constraints reduce A type's leverage causing the return volatility to decrease relative to the frictionless benchmark.

The right panel of Figure 5 presents intermediary leverage in the equilibrium with a Value-at-Risk-type state-dependent margin constraint. In the model with margin constraints, financial sector and banks both exhibit *countercyclical* leverage, while broker-dealers could have procyclical leverage when constraints bind. This is again consistent with the evidence presented in Figure 1a (dashed red line for holding company leverage and solid blue line for broker-dealer leverage) and also recently documented in the empirical intermediary asset pricing literature in AEM and HKM.

6.3 Heterogeneous versus Representative Intermediaries

Since the aggregate endowment in equation (1) is i.i.d. with constant volatility, variation in risk premia is only due to wealth distributions and intermediation frictions captured by state variables x and y. State variable x, representing the wealth share of the aggregate financial sector in the economy, is the primary determinant of time-varying risk premia in representative intermediary models. In my model with a heterogeneous financial sector, however, the wealth distribution among intermediaries (captured by state variable y) also contributes to the variation in risk premia. In this section, I answer the following question: What fraction of the variation in risk premia can be attributed to the state variable y, a measure of the composition of the financial sector?

To answer this question and investigate the role of heterogeneity in the financial sector, I compare the main results of the three-sector model from above with the ones from an economy with identical fundamentals but *two* heterogeneous agents instead: a household sector (C types identical to the main model) and a *representative* intermediary sector (I), where I agents face the same endogenous margin constraints as in the original model (equation 7 with $\nu = 1$).

Most of the parameters in the two-agent representative intermediary model are identical to the ones in the main model listed in Table 1: household's risk aversion $\gamma_C = 15$, EIS for household and the representative intermediary $\psi_C = \psi_I = 1.5$, rate of time preference $\rho = .001$, growth rate and volatility of the aggregate endowment $\mu_D = .022$, $\sigma_D = .035$, and agents birth and death rates $\kappa = .0154$ (exactly as in the original three-sector model). I set risk aversion of the representative intermediary sector to $\gamma_I = 3.3$: the wealth-weighted "harmonic average" of relative risk aversion for the financial sector (A and B intermediaries) in the main model (with heterogeneous intermediaries) evaluated at the stochastic steady state.[36] Population share of the representative intermediary sector is set to $\bar{x} = .12$: sum of population shares of two intermediary sectors in the main model, $\bar{u}$ and $\bar{v}$ from Table 1. Note that with a representative intermediary, there is only *one* endogenous state variable x: the wealth share of the intermediary sector.[37]

I simulate representative- and heterogeneous-intermediary models for 3,000 quarters 20,000 times and examine the distribution of risk premium volatility.[38] As expected, the model with heterogeneous intermediaries exhibits more variation in risk premia than the one with a representative financial sector. Since the aggregate intermediary sectors in both models are (almost) identical,

[36]The wealth-weighted harmonic mean of risk aversions for the intermediary sector in main three-agent model is $\gamma_I = \left(\frac{y}{\gamma_A} + \frac{1-y}{\gamma_B}\right)^{-1}$. Using values for $\gamma_A = 2.5$ and $\gamma_B = 5.5$ from Table 1, at the stochastic steady state $y_{ss} = 0.56$, we get $\gamma_{I,ss} = 3.3$.

[37]The wealth share of the financial sector is the key state variable in existing models with a representative intermediary sector. See He and Krishnamurthy (2013), Brunnermeier and Sannikov (2014), Di Tella (2017), and Drechsler et al. (2018), for example.

[38]Figure G.1 in Appendix C shows these distributions.

any excess variation in risk premia in the heterogeneous intermediary model has to be due to state variable y. In my calibration, approximately 20% of the variation in risk premia can be attributed to heterogeneity in the financial sector (state variable y).[39] Therefore, failing to account for heterogeneity among intermediaries can lead to missing a substantial portion of the variation in risk premia.

6.4 Implications of Financial Sector's Balance Sheet Adjustments

As discussed in Section 2 and documented in Figures 1a and 1b, during the height of the financial crisis, broker-dealers substantially delevered (reduced leverage by approximately 47% relative the previous quarter), while during the same period, holding companies increased leverage by 72%. In this section, I measure the impact of this balance sheet adjustments within intermediaries on risk premia and endogenous risk.

When the least risk averse A types face binding margin constraints, they are forced to reduce leverage by selling assets. To clear the risky asset market, the more risk averse agents (B and C types) need to take on a larger portion of the asset than they would in the absence of constraints. In order to entice them to buy more, the risk premium must increase.

As mentioned the model has difficulties matching the high level and substantial variation of intermediary leverage during the Great Recession. To study the impact of balance sheet adjustments within the financial sector, I perform the following exercise: I try to match the aforementioned increase and decrease in leverage of broker-dealers and holding companies, respectively, by tightening the margin constraint faced by A types in baseline calibration in Table 1. This is consistent with empirical evidence that the contraction in repo market financing during the recent financial crisis hit broker-dealers (represented by A agents in the model) particularly hard and forced them to deleverage (see Gorton and Metrick, 2012 and He et al., 2010, for example). As noted earlier in Section 6.1, tighter margin constraints lead to a decrease in total leverage and thus a *lower* volatility, which is counterfactual. To achieve an increase in volatility consistent with the empirical

[39]Notice that the horizontal axis in Figure G.1 is the standard deviation of the risk premium for the endowment claim, which has relatively low volatility (3.5% in my calibration) relative to the market (approximately 16%). Therefore, equity premium volatility implied by the model is about five to six times larger than that of the endowment claim (approximately 1% for the heterogeneous intermediary model, for example).

evidence, I also make households relatively more risk averse than intermediaries.

Figure 6 presents results of this exercise. Tighter margin constraints is implemented by an increase in the parameter α. I also increases risk aversion of households relative to the intermediary sector (lower γ_I/γ_C) to obtain an increase in volatility. The top left panel shows that at model's stochastic steady state, increasing parameter α by 50% (consistent with the rise in repo-haircuts index during the 2008 crisis from Gorton and Metrick (2012)'s Fig. 4) and reducing γ_I/γ_C by 26% (from 0.61 to 0.45), results in approximately 48% decline in A types' leverage. This deleveraging is very close to what broker-dealers experienced during the crisis.

To clear the risky asset market, the risk premium must increase enough to entice more risk averse agents to take on a larger supply of the asset than before the shock. More risk averse holding companies increase leverage as a result of dealers' deleveraging (the top right and middle left panels): B type leverage rises by 79% and households' holding of the risky asset remain relatively unchanged (see the middle left panel of Figure 6). As the middle right and bottom left panels of Figure 6 show, this dealer deleveraging results in an approximately 55% increase in the risk premium and Sharpe ratio a 5% rise in endogenous volatility.

More importantly, asset reallocations between intermediaries do not impact the aggregate wealth share of the financial sector, and thus do no affect risk premia and volatility in a model featuring representative intermediaries. My model with a heterogeneous financial sector captures the implications of these balance sheet adjustments for equilibrium objects.

6.5 Empirical Predictions of the Model

In this section I present two main empirical predictions of the model which I test in the data: (1) The composition of the financial sector, captured in the state variable y, negatively predicts returns, and (2) shocks to this measure of composition is priced in the cross-section of assets with a *positive* price of risk

State variables load *Positively* on aggregate shocks. Figure 7 presents the diffusions of state variables x and y (σ_x and σ_y in equation 16 and 17) in the equilibrium with state-dependent margin constraint. From equation 16, σ_x always remains positive. Since, in my calibration, σ_y is also positive in the entire state space, both state variables are *procyclical*: following a negative dZ shock, both state variables go down.

The empirical prediction of the model is that state variables *negatively* predict future excess returns. In particular, times when the more aggressive intermediaries are relatively more impaired in the financial sector (i.e., when state variable y is low), coincide with high marginal utility states when the risk premia is high.

Price of Risk is *decreasing* in x and y. The left (right) panel of Figure 8 plots the Sharpe ratio of the risky asset as a function of state variable x (y) for three different values of the other state variable: Its unconditional mean, and its unconditional mean ± 3 standard deviations. We see that the Sharpe ratio is *decreasing* in both sate variables x and y. In Assets that pay

The empirical prediction of the model is that shocks to state variables are priced in the cross-section of expected returns with a *positive* price of risk. In particular, asset that pay well when the more aggressive intermediaries are relatively more impaired in the financial sector (i.e., when state variable y is low), are valuable hedges and demand *lower* excess returns.

6.6 Reconciling Empirical Evidence in AEM and HKM

In this section, I argue that my model featuring heterogeneous intermediaries and leverage constraints can reconcile the conflicting cross-sectional asset pricing evidence, recently documented in AEM and HKM.[40] As mentioned earlier, AEM and HKM find opposite signs for the price of intermediary leverage shocks in the cross-section of asset returns and thus conflicting cyclical dynamics of the intermediary leverage.[41]

Importantly, in reaching these seemingly contradictory results, AEM and HKM measure intermediary leverage in different parts of the financial sector: broker-dealers and bank holding

[40] He et al. (2017) present a simple, one-period model in their appendix which was originally suggested by Alexi Savov in a conference discussion of the paper. The model can reconcile the contradiction between HKM's results and the ones documented in AEM. This static framework, however, is unable to capture implications of balance sheet adjustments within the financial sector for risk premium, the price of risk, and volatility discussed above.

[41] As mentioned above, the data for security brokers-dealers are from Table L.130 of the Financial Accounts of the United States (Flow of Funds). The underlying source for this data comes from FOCUS and FOGS quarterly reports filed with the SEC by these broker-dealers *in isolation* from other parts of their holding companies which are not publicly available. Data for publicly-traded holding companies of primary dealers are from CRSP/Compustat and Datastream. For a more detailed description of data sources, see Appendix E.

companies, respectively. AEM use shocks to book leverage of broker-dealers to construct an intermediary stochastic discount factor (SDF) and show that it prices equity and bond portfolios with a *positive* price of risk implying *procyclical* intermediary leverage. HKM, on the other hand, find that shocks to leverage of bank holding companies for the New York Fed's primary dealers price the cross-section of returns for many asset classes with a *negative* price of risk. In contrast to AEM, HKM's negative price of risk suggests that intermediary leverage is *countercyclical*.

A direct implication of the opposite flows and leverage dynamics within the financial sector in my model with heterogeneous intermediaries is that measuring the price of risk for shocks to leverage of *different* intermediaries will result in opposite signs. Thus, these seemingly contradictory asset pricing results can be reconciled in a model where the intermediary sector is modeled as two heterogeneous sectors facing financial constraints. Moreover, in Section 7, I will show that the composition of the financial sector plays an important role for time-series predictability and also has significant explanatory power for the cross-section of expected returns.[42]

7 Empirical Implications of the Model

In this section, I study empirical implications of the model discussed in Section 6.5 and show that the composition of the intermediary sector, measured by the wealth share of broker-dealers in the financial sector, matters for time-series return predictability and is also priced in the cross-section of expected returns. I focus on asset pricing implications of the model with time-varying margin constraints ($\nu = 1$ in equation 7). I show that consistent with model's predictions, an empirical measure of the state variable y, which measures the composition of the financial sector, has strong forecasting power for returns on various asset classes beyond factors already established in the literature to predict returns. Moreover, innovation in this wealth distribution prices the cross-section of equity and bond portfolios.

[42] It is important to note that the intermediary leverage is endogenous, and the fact that shocks to leverage are priced does not necessarily mean that intermediaries are the marginal investors. It could well mean that leverage is proxy for aggregate risk aversion. See Santos and Veronesi (2016) and Haddad and Muir (2018) for more details.

7.1 Measuring Heterogeneity in the Intermediary Sector

As mentioned earlier, potentially there are other ways to construct model's two state variables. A valid way to construct them in the data is by measuring the wealth share of the financial sector in the aggregate economy (for x) and that of BDs in the financial sector (for y). In many recent representative intermediary-based models (e.g., He and Krishnamurthy, 2013 and Brunnermeier and Sannikov, 2014), x is the main state variable. Moreover, with this choice of y, in the absence of heterogeneous intermediaries, my model collapses to one with x as the only state variable similar to the recent intermediary-based models. Thus, I believe this presents a natural way to construct the state variables of my model.

The price of risk in my model is time-varying and depends on wealth share of the financial sector, state variable x, as well as, broker-dealers' wealth share in the intermediary sector, state variable y in the model. Since financial sector's wealth share, x, is the key state variable in many existing models with a *representative* intermediary sector, in this section, I emphasize the importance of wealth distribution within the intermediary sector, captured by state variable y, for forecasting future returns.[43]

Since risk premium is decreasing in y (see Figure 8 in the Appendix), an asset that pays well when y is low is less risky. Thus, my model predicts that a higher wealth share for BDs in the financial sector forecasts higher prices and thus, *negatively* predicts future returns.

In the data, I compute wealth share of the financial sector as the ratio of their market equity of to the total market value of firms in the CRSP universe:

$$x_t^{data} = \frac{\text{Market capitalization of the financial sector}_t}{\text{Total market capitalization of the CRSP universe}_t}. \tag{29}$$

I use monthly equity data from CRSP to compute x^{data}. The financial sector is identified as firms in the CRSP universe for whom the first two digits of the header standard industry classification (SIC) code equals 60 through 67.[44]

[43] Adrian, Moench, and Shin (2014b) study return predictability in representative intermediary models and show book leverage of broker-dealers negatively forecasts future equity and bond returns. He et al. (2017) also run time-series predictive regressions and show the squared reciprocal of capital ratio for bank holding companies of NY Fed's primary dealers positively predicts future returns for many asset classes.

[44] This definition of the financial sector has been commonly used in the literature. See Giglio, Kelly, and Pruitt

Since I don't have access to market data for broker-dealers, model's second state variable, y (wealth share of broker-dealers in the financial sector), is computed as the ratio of BD's book equity to the sum of commercial bank and BD book equities from the Flow of Funds Tables L.110 and L.130, respectively:

$$y_t^{data} = \frac{\text{Book equity of BDs}_t}{\text{Book equity of Commercial Banks}_t + \text{Book equity of BDs}_t}, \tag{30}$$

where equity is computed by subtracting total liabilities (excluding miscellaneous liabilities) from (mark-to-market) total financial assets.[45] For a detailed description of the data, see Appendix E.

As mentioned earlier, A agents in the model are not meant only to represent broker-dealers (BDs). BDs are good proxies for a set of intermediaries that face tightening leverage constraints in bad times, and their actions resemble these agents in the model (e.g., hedge funds). Moreover, BDs are intermediaries for which I have access to aggregate balance sheet data from the Flow of Funds. Unfortunately, I cannot measure the hedge fund sector as well as I can do BDs. Similarly, B agents are meant to describe intermediaries who face equity capital constraints (e.g., commercial banks and insurance companies). Although this simple classification ignores potentially important institutional details, it goes one step further than the existing literature to study the asset pricing of implications of heterogeneous intermediaries.

It is also important to point out that many of the primary broker-dealers are subsidiaries of large US bank holding companies, If the internal capital markets represent an important source of funding for these broker-dealer subsidiaries, it may seem unnecessary to treat BDs separate form BHCs. For instance, if the commercial banking affiliate of a BHC experiences large deposit inflows during a crisis (as documented in Gatev and Strahan (2006), for example), the access to internal capital markets may greatly mitigate the financial distress in the broker-dealer subsidiary.

(2016), for example.

[45] In unreported results, using 49 Fama-French industry definitions, I identify publicly-traded broker-dealers as all US firms in the CRSP universe with standard industry classification (SIC) codes 6211 (Security brokers, dealers & flotation companies) or 6221 (commodity contracts brokers & dealers). I then equivalently define state variable y using *market* data as $y_t^{data,mkt} = \frac{\text{Market cap of dealers}_t}{\text{Market cap of the financial sector}_t}$. During the sample period (1971Q1-2010Q4), time series of $y^{data,mkt}$ is highly positively correlated with y^{data} computed from book values (in equation 30) with correlation coefficient of 0.55 (t-stat of 8.88). Prior to 2010, book and market series are even more strongly positively correlated (correlation coefficient of 0.68 with t-stat of 11.74). Post 2010, however, the two series become negatively correlated with coefficient of -0.5 (t-stat of -2.92).

Federal Reserve's Regulation W prohibits BHCs from freely using deposits in their commercial banking arm to fund the broker-dealer subsidiary when repo markets collapse.[46] Moreover, Gupta (2018) documents that the internal capital markets are not frictionless and do not behave very differently from the external ones during the financial crisis. He shows that although the aggregate size of the dealer internal capital markets quadrupled from $300 billion in 2001 to $1.2 trillion by 2007, these inter-affiliate repo and securities loans collapsed by 37% in 2008.

Table 2 reports the mean, standard deviation, and autocorrelation of the state variables in the data, both in levels and innovations. The factors are autocorrelated in levels but not in changes.

Figure 9 shows the time-series of x^{data} and y^{data} using the CRSP and Flow of Funds data confirming the dramatic growth of the sector from 1980 to the onset of the recent financial crisis. Consistent with the model, x^{data} and y^{data} are both *procyclical*: innovations in x^{data} and y^{data} are both *positively* correlated with the innovations in the real GDP with correlation coefficients of 0.24 (t-stat of 2.05) and 0.21 (t-stat of 2.99), respectively.[47]

7.2 Intermediary Heterogeneity and Time-Series Predictability

The risk premium in my model is time-varying due to its association with wealth share of the financial sector, state variable x, as well as, broker-dealers' wealth share in the intermediary sector, state variable y in the model. As a result, expected returns are time-varying in the model and are predictable using lagged state variables as predictors. Since financial sector's wealth share, x, is the key state variable in existing models with a *representative* intermediary sector, in this section, I emphasize the importance of wealth distribution within the intermediary sector, captured by state variable y, for forecasting future returns.[48]

[46] Sections 23A and 23B of Federal Reserve's Regulation W set limits on the amount of covered transactions between a bank and its affiliates and require these transactions to be collateralized and on market terms and conditions. Covered transactions include loans and other extensions of credit to an affiliate, investments in the securities of a subsidiary, purchases of assets from an affiliate, and certain other transactions that expose the bank to the risks of its affiliates. See Appendix D for more details on Regulation W.

[47] It is worth pointing out that the decline in y^{data} post 2009 is because two of the largest broker-dealers (Goldman Sachs and Morgan Stanley) became bank holding companies in 2009Q1. Two other broker dealers were also acquired by bank holding companies: J.P. Morgan purchased Bear Sterns and Merrill Lynch became part of Bank of America.

[48] Adrian et al. (2014b) study return predictability in representative intermediary models and show book leverage of broker-dealers negatively forecasts future equity and bond returns. He et al. (2017) also run time-series predictive regressions and show the squared reciprocal of capital ratio for bank holding companies of NY Fed's primary dealers positively predicts future returns for many asset classes.

Due to the presence of a single aggregate dividend shock, the two state variables are positively correlated in the model. As a result, one might be concerned about multicollinearity when both x and y are included as forecasting variables in predictive regressions. However, the model with occasionally binding margin constraints exhibits highly nonlinear dynamics. In particular, running predictive regressions *unconditionally* using very long sample of simulated data in the model (I used 300-year long sample and 20,000 simulations), results in negative and significant coefficients on *both* state variables x and y.

Another concern in the empirical results that follows may be the presence of time-series trend in the state variable y in Figure 9. In Appendix F, I consider various ways to detrend the state variables and perform additional robustness checks with the cyclical components. The main predictive regression results, which use the levels of state variables as predictors, are also robust to using the detrended variables.

As discussed above, state variable y is procyclical (see Figure 7 in Appendix C.2): times when the more aggressive intermediary is relatively more impaired in the financial sector, i.e. when y is low, coincide with high marginal utility states where the risk premia is high. As such, my model predicts that a higher wealth share for BDs in the financial sector forecasts higher prices (lower returns) thus, *negatively* predicting future returns. To test this hypothesis, I regress one-year-ahead holding period excess return (from quarter $t+1$ to $t+4$) for asset i on the lagged wealth share of BDs in the financial sector, y_{data} defined in equation (30), and controls:

$$R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y \, y_t + \gamma^i_{\text{Ctrl}} \, \text{Ctrl}_t + \varepsilon^i_{t+1\to t+4} \tag{31}$$

where $R^i - r^f$ is the average excess return on asset i, and Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: wealth share of the aggregate financial sector (x from the model), fluctuations in the aggregate consumption-wealth ratio (*cay* variable) defined in Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev, Tauchen, and Zhou (2009).

The model predicts a *negative* and significant coefficients γ^i_y: times when wealth share of broker-dealers in the financial sector is high are associated with low marginal utility states where asset prices are high and future expected returns are low. As test assets, I use value- and equally-weighted CRSP portfolios, mean excess return of 25 size/book-to-market and 10 momentum portfolios from Ken French's data library, as well as, an equally-weighted portfolio of assets within each non-equity class studied in HKM available form Asaf Manela's website.[49] The sample is quarterly starts in 1970Q1 and ends in 2017Q4 for equity portfolios and in 2012Q4 for non-equity assets (limited by data availability).

Table 3 presents results of the univariate predictive regressions in (31), with state variable y as the only predictor, for different test assets mentioned above. In Column (1), where the dependent variable is the market excess return, I can directly verify model's prediction that state variable y should negatively predict future aggregate risk premia: an increase in the measure of the composition of the financial sector (state variable y) of 1 percentage point in deviation from its mean decreases the expected excess return by 1.16 percentage points (per quarter). Consistent with model predictions, we observe negative and significant $\hat{\gamma}_y$ for Market, size/book-to-market, momentum, sovereign bonds, and options portfolios. For most asset classes $\hat{\gamma}_y$ is negative, as expected: this measure of the composition of the financial sector *negatively* forecast future returns.[50]

To examine whether forecasting relationships are stable over time, and an investor could have profited from observing the predictor variable y, I follow Goyal and Welch (2008) and Campbell and Thompson (2008) to evaluate the out-of-sample performance of the predictive regressions. I compute an out-of-sample R^2 statistic (R^2_{OOS}) as:

$$R^2_{\text{OOS}} = 1 - \frac{\sum_{t=1}^{T} (r_t - \hat{r}_t)^2}{\sum_{t=1}^{T} (r_t - \bar{r}_t)^2},$$

where $\hat{r}_t$ is the fitted value from a predictive regression estimated through period $t-1$, and $\bar{r}_t$ is the historical average return estimated through period $t-1$. The R^2_{OOS} for market excess

[49]Non-equity assets in HKM are mostly from previous studies. See Appendix E and He et al. (2017) for more details on test assets.

[50]The predictor variable y_t is very persistent with AR(1) coefficient around 0.96 in quarterly data. I verify (in unreported regressions) that the absolute value of the regression coefficient $\hat{\gamma}_y$ and the R^2 both rise with the forecast horizon (see Cochrane (2005)'s Chapter 20 for more details). The estimates in Table 3 and Table G.3 in the Appendix are corrected for the Stambaugh (1999) bias. Moreover, if I use growth rate of variable y (I used one- and five-year growth rates) as predictors, I still observe negative and significant coefficients γ_y.

return, Column (1) of Table 3, is approximately 4%. To assess the economic significance of return predictability, I use Campbell and Thompson (2008)'s simple metric: the increase in expected returns of a one-period mean-variance (MV) investor from observing the predictor variable y. A quarterly out-of-sample R^2 of 4% leads to an increase in expected returns of approximately 3% per year for a MV agents with a risk-aversion coefficient of 5.

Table 4 provides results of the predictive regression in (31) adding several control variables mentioned above to regressions in Table 3.[51] In Column (1), as a benchmark, I report the forecasting regression for the market risk premium using only aforementioned control variables as predictors. In Column (2), I add dealer wealth share in the financial sector, y, as an additional predictor. Comparing Columns (1) and (2), it is particularly important to point out that the composition of the financial sector, captured in wealth distribution y, leads an additional 15 percentage points predictive power for future market excess returns over variables already known in the literature to forecast returns: the R^2 of the predictive regression on the market excess return goes from 0.28 to 0.43 when y is included in the regression in Column (2) of Table 4.[52] It similarly reports negative and significant $\hat{\gamma}_y$ for Market, size/book-to-market, momentum, sovereign bonds, and options portfolios.

In Appendix F, I provide series of additional robustness checks for the time-series predictability regressions above. I first show that the predictive regression results are robust to excluding the Great Recession (years 2007 to 2009) from the sample. So, it is not just the financial crisis that drives this predictability results. As mentioned above, one also might be worried about time-series trends in state variable y. I show that the main predictive regressions in Table 4 are robust to using the cyclical component of state variable y in the data. I also rerun the forecasting regression in Table 4 by adding AEM's broker-dealer leverage ratio, HKM's intermediary capital ratio for different asset classes. Adding these additional predictors, however, does not change the sign and significance of the coefficient $\hat{\gamma}_y$.

In summary, in this section, I provided strong empirical evidence that the composition of the

[51] The sample is now shorter and starts in 1990Q1 due to data availability for Bollerslev et al. (2009)'s variance risk premium, one of the control variables used in return forecasting regression in equation (31).

[52] In one-quarter ahead predictive regressions, the incremental R^2 increases by 5 percentages points (from 0.22 to 0.27) when intermediary composition variable y is added to the regression.

financial sector, captured by state variable y, matters for prices, beyond the health of the aggregate financial sector: it has strong predictive power for excess return on many assets beyond variables from *representative* intermediary asset pricing models, as well as, the ones already known in the literature to predict return.

7.3 Intermediary Heterogeneity and the Cross-Section of Asset Returns

As mentioned above, the model with heterogeneous financial intermediaries can reconcile seemingly contradictory evidence for the sign of estimated price of risk for intermediary leverage shocks documented in AEM and HKM. In this section, I explore the implications of a heterogeneous financial sector for the cross-section of returns.

As shown in Figure 2 (and also in the bottom right panel of Figure G.3 in Appendix C.2), risk premium on the endowment claim is *decreasing* in both state variables x and y.[53] This suggests assets that pay poorly when: (i) the financial sector is less capitalized (i.e. when x is low), and/or (ii) wealth share of broker-dealers in the financial sector is small (i.e. when y is low), are riskier and should command higher expected returns. I emphasize that point (ii) can only be made in a model with heterogeneous financial intermediaries.

7.3.1 Cross-sectional Asset Pricing Tests

Similar to He et al. (2017), I construct the growth rate to dealers' wealth share in the financial sector, denoted y_t^{Δ}, as follows. I estimate a shock to dealer wealth share in levels, ς_t, as an AR(1) innovation in the regression: $y_t = \phi_0 + \phi y_{t-1} + \eta_t$. I then convert these innovations to a growth rate by dividing them by the lagged wealth share:

$$\text{HIFac} = y_t^{\Delta} = \frac{\eta_t}{y_{t-1}} \tag{32}$$

I call this wealth share growth rate the *heterogeneous-intermediary factor* (HIFac) and use it to perform cross-sectional asset pricing tests. For each asset i, I first estimate betas from time-series

[53] This is true even in the absence of margin constraints as shown in the bottom right panel of Figure G.2.

regressions of portfolio excess returns on the risk factors:

$$R^e_{i,t} = a_i + \beta'_{i,f}\mathbf{f}_t + \vartheta_{i,t}, \quad i = 1, \dots, N, \tag{33}$$

where $\mathbf{f}$ represents the $K \times 1$ vector of risk factors. I consider four cases: (1) $\mathbf{f}_t = \text{HIFac}_t$, (2) $\mathbf{f}_t = [\text{HIFac}_t \quad \text{MktRF}_t]'$, (3) $\mathbf{f}_t = [\text{HIFac}_t \quad \text{AEM}_t]'$, and (4) $\mathbf{f}_t = [\text{HIFac}_t \quad \text{HKM}_t \quad \text{MktRF}_t]'$, where AEM is the broker-dealer leverage factor from Adrian et al. (2014a), HKM is the intermediary capital risk factor from He et al. (2017), and MktRF represents the market risk premium. For comparison, I also report the pricing performance of AEM and HKM factors.

Next, in order to estimate factor risk prices, λ_f, I run a cross-sectional regression of average excess returns on the estimated risk exposures $\hat{\beta}_{i,f}$:

$$\mathbb{E}\left[R^e_{i,t}\right] = \alpha_i + \hat{\beta}'_{i,f}\lambda_f + \zeta_i, \quad i = 1, \dots, N, \tag{34}$$

As mentioned above, the model predicts a *positive* and significant sign for the estimated price of risk λ_{HiFac}.

I test the ability of the heterogeneous intermediary factor in pricing the cross-section of 55 equity and bond portfolios: the test assets are 25 size and book-to-market and 10 momentum portfolios from Ken French's website, 10 maturity-sorted US government bond portfolios from CRSP's Fama Bond dataset with maturities up to five years in six month intervals, and 10 US corporate bond portfolios sorted on yield spreads from Nozawa (2017) obtained from Asaf Manela's website. I choose equity and bond portfolios as test assets due to the availability of longer time-series than others such as options and CDS. Since I use many test assets beyond size and book-to-market portfolios, my model avoids the typical criticisms of asset pricing tests discussed in Lewellen, Nagel, and Shanken (2010).

Table 5 presents the main asset pricing results. Below estimated risk prices I report Shanken (1992) t-statistics that corrects for estimation error in betas and cross-correlations. I also report Fama and MacBeth (1973) t-statistics by running period-by-period cross-sectional regressions and computing standard errors of the time-series average of λs. I report cross-sectional R^2 and the mean absolute pricing error (MAPE), calculated as $\frac{1}{N}\sum|\zeta|$ where N is the number of test assets, as measures of model fit. I also report a $\chi^2(N-K)$ statistic (K is te number of factors) that tests if the pricing errors are jointly zero.

Column 1 of Table 5 reports the results of heterogeneous intermediary factor as a single pricing factor. The estimated price of risk is positive, which means assets that pay well in states with a low broker-dealer wealth share in the financial sector (i.e. assets with low betas on y_t) are valuable hedges and have lower expected returns in equilibrium. This risk price estimate confirms the procyclicality of broker-dealer wealth share y_t documented in Figure 9. The adjusted R^2 is 61% while the total MAPE is only 1.86%. The single-factor model can explains 62% of the variation in average returns in these cross-sections, with an average absolute pricing error around 1.8% per annum. Figure G.5 in the Appendix, visually shows the HIFac's pricing performance in the cross-section of equity and bond returns by plotting realized versus predicted returns.

For robustness and comparison with recent empirical work, in Columns 2–6, I add additional pricing factors. In Column 2, I include market risk premium, MktRF, as an additional factor. However, the price of risk for MktRF is not statistically significant and in terms of almost all test statistics, the two-factor model is nearly identical to the single-factor model in Column 1. The market adds essentially no explanatory power to the intermediary heterogeneity factor.

In Columns 3 and 5, for reference, I present performances of the pricing factors in AEM and HKM, respectively. AEM use a leverage factor defined as the seasonally adjusted growth rate in broker–dealer book leverage level from Flow of Funds. A shown in Column 3, for the test assets mentioned above, HIFac outperforms AEM with 38% lower MAPE (1.83% versus AEM's 2.96%) and 56% higher cross-sectional R^2 (61% compared to 39% in AEM).[54]

In HKM, the pricing factors are the market risk premium (MktRF) and shocks to intermediary capital ratio defined as the ratio of total market equity to total market assets (book debt plus

[54] The pricing performance of AEM reported in their Table III and shown in Figure 1 of their paper, is substantially better than the one reported in Column 3 of Table 5 (their reported MAPE is only 1.31% and $R^2 = 0.77$). This difference can stem from two possible sources: (i) In 2015, the Federal Reserve substantially revised and updated Flow of Funds historical data for security broker-dealers, changing the way assets and liabilities were counted. They specifically changed their handling of using gross vs net repo. For more detail, see Z.1 Financial Accounts Technical Q&As. (ii) The test assets used in this paper are different from AEM's. I have the same 35 equity portfolios (25 size/book-to-market and 10 momentum portfolios) but use both Treasury and corporate bonds from Nozawa (2017), while AEM only use 6 Treasury bonds sorted by maturity from CRSP.

market equity) for New York Fed's primary dealer holding companies. A shown in Column 5, my model with a single pricing factor performs almost as well as HKM's two factor model with nearly identical MAPE (1.83% vs. 1.89% for HKM) and cross-sectional R^2 (61% vs. HKM's 63%).

In Column 4, I add leverage factor from AEM to evaluate a model with two pricing factors: HIFac and AEM. Addition of AEM's leverage factor does not make price of HIFac risk insignificant or change its sign. This even raises the cross-sectional R^2 to 72%. Finally, in Column 6, I add two pricing factors from HKM: MktRF and shocks to intermediary capital ratio. Again, λ_{HIFac} remains positive and significant. Note that since HIFac is positively corrected with both HKM and AEM factors, it is not surprising that λ_{HIFac} has weaker statistical significance in the presence of these additional factors.[55]

In summary, The results in Table 5 demonstrates that heterogeneity in the financial sectors has explanatory power for the cross-section of expected returns even in the presence of representative intermediary asset pricing factors presented in AEM and HKM.

7.3.2 Sorted Portfolios on Exposures to Heterogeneous Intermediary Factor

The positive price of risk associated with shocks to wealth share of dealers in the financial sector means assets that pay more in states with a low dealer wealth share (i.e. assets with low betas on y_t shocks) are viewed as hedges thus have lower expected returns in equilibrium.

To empirically verify the positive price of risk for innovations in the wealth share of dealers in the financial sector, I sort stocks based on their exposures to these shocks and form portfolios by quintiles on a 10-year trailing window. I consider all common stocks (share codes 10 and 11) in the CRSP universe from Amex, NASDAQ, and NYSE (exchange codes 1,2, and 3). For every stock i at quarter t, I regress its quarterly excess return on constant and innovations in the heterogeneous intermediary factor (HIFac), defined in equation (32):

$$R^e_{i,t} = \alpha_i + \beta_{i,\text{HIFac}}\,\text{HIFac}_t + \xi_{i,t} \tag{35}$$

The coefficient $\beta_{i,\text{HIFac}}$ measures the exposure of firm i's stock to the factor's innovations. I then

[55] HIFac has positive correlation of 13% and 9% with AEM's leverage and HKM's capital risk factors, respectively.

sort stocks into quintiles every quarter according to their $\beta_{i,\text{HIFac}}$.

Consistent with model's implications, when sorted on β_{HIFac}, average risk premia are increasing from the portfolio of low-beta stocks to the high-beta quintile. In the Appendix, I report the average returns of the beta-sorted portfolios in Table G.4, along with return volatilities, average book-to-market ratio, average market cap, and alphas from CAPM and Fama-French three-factor models. Excess returns are monotonically increasing from quintile one to five and the top portfolio earns an approximately 5% premium over the lowest quintile. In Appendix F, I further verify the results above are robust to double-sorting with asset pricing factors from recent models with representative intermediaries.

This exercise demonstrates that the heterogeneity in the financial sector is an important risk factor and has pricing information above and beyond representative intermediary asset pricing factors in AEM and HKM: even within portfolios sorted based on AEM or HKM factor betas, I see a monotonic progression in returns from low- to high-HIFac beta portfolios.

7.4 Additional Robustness Checks

In Appendix F, I project the heterogeneous intermediary factor (HIFac) onto the space of traded returns to form a factor-mimicking portfolio that mimics the HIFac. To further verify that this heterogeneity an important source of risk, I evaluate the heterogeneous intermediary factor-mimicking portfolio (HIMP) relative to the mimicking portfolios for representative intermediary factors in AEM and HKM. I show that the mimicking portfolios for these representative intermediary factors cannot fully span the HIMP and there is more to be captured by the heterogeneity within the financial sector. This exercise confirms my earlier results: the heterogeneity in the financial sector is an important risk factor and has pricing information above and beyond representative intermediary asset pricing factors in AEM and HKM.

8 Conclusion

This paper studies the asset pricing implications of heterogeneity among financial intermediaries. Evidence on large balance sheet adjustments within the intermediary sector during the Great

Recession is at odds with existing models featuring representative intermediaries. To explain and study the implications of these massive asset reallocations within the financial sector, in this paper, I present a model with two main ingredients: intermediaries heterogeneous in their aggressiveness, and occasionally binding leverage constraints. Heterogeneity in intermediaries' risk appetite and margin constraints are both necessary in matching and understanding reallocations within the financial sector.

My model implies that a dealer deleveraging episode, comparable in magnitudes to the one observed during the recent financial crisis, leads to an approximately 55% increase in the risk premia and a 5% increase in endogenous volatility. In contrast, since balance sheet adjustments among different intermediaries does not affect the wealth share of the aggregate financial sector, in models with representative intermediaries these asset reallocations have no impact on asset prices and risk premia.

I show that the composition of the financial sector accounts for a substantial portion of the variation in risk premia in the model. With an independent and identically distributed aggregate endowment, the variation in expected returns is entirely due to intermediation frictions captured by both the health of the overall financial sector as well as its composition. I show that the wealth distribution among intermediaries, a measure of the composition of the financial sector, accounts for approximately 20% incremental variation in risk premia over a representative intermediary model.

The model also generates opposite cyclical dynamics for leverage of the two intermediary sectors, reconciling empirical evidence that has previously seemed contradictory through the lens of representative intermediary asset pricing models of AEM and HKM. To construct the SDF, AEM and HKM measure marginal utilities of different financial intermediaries: security broker-dealers, and bank holding companies, respectively. Given the economic mechanism of my model, it does not seem surprising that they arrive at conflicting asset pricing results.

I examine the empirical implications of the model for time-series predictability and the cross-section of returns. Consistent with the model, wealth share of broker-dealers in the financial sector, a measure of heterogeneity among intermediaries, strongly and negatively forecasts future excess returns on many assets. In particular, it leads to additional predictive power for market risk premium beyond many established forecasting variables in the literature. I also show that using only shocks to the relative wealth share of broker-dealers in the financial sector, explains the cross-section of equity and bond returns about as well or better than existing intermediary asset pricing models. I further document that including aggregate intermediary leverage as a second asset pricing factor increases cross-sectional fit by at least 10 percentages points, depending on whether the factor is the leverage of broker-dealers or bank holding companies.

References

Adrian, Tobias, and Nina Boyarchenko, 2015, Intermediary leverage cycles and financial stability Working Paper, Federal Reserve Bank of New York Staff Reports.

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014a, Financial intermediaries and the cross-section of asset returns, *The Journal of Finance* 69, 2557–2596.

Adrian, Tobias, Emanuel Moench, and Hyun Song Shin, 2014b, Dynamic leverage asset pricing Federal Reserve Bank of New York Staff Report No. 625.

Adrian, Tobias, and Hyun Song Shin, 2014, Procyclical leverage and value-at-risk, *Review of Financial Studies* 27, 373–403.

Ang, Andrew, Sergiy Gorovyy, and Gregory B. van Inwegen, 2011, Hedge fund leverage, *Journal of Financial Economics* 102, 102 – 126.

Avraham, Dafna, Patricia Selvaggi, and James Vickery, 2012, A structural view of U.S. bank holding companies, *FRBNY Economic Policy Review* 18, 65–81.

Bansal, Ravi, Dana Kiku, and Amir Yaron, 2012, An empirical evaluation of the long-run risks model for asset prices, *Critical Finance Review* 1, 183–221.

Bansal, Ravi, and Ivan Shaliastovich, 2013, A long-run risks explanation of predictability puzzles in bond and currency markets, *The Review of Financial Studies* 26, 1–33.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The Journal of Finance* 59, 1481–1509.

Basak, Suleyman, and Domenico Cuoco, 1998, An equilibrium model with restricted stock market participation, *The Review of Financial Studies* 11, 309–341.

Begenau, Juliane, Saki Bigio, Jeremy Majerovitz, and Matias Vieyra, 2019, Banks adjust slowly: Evidence and lessons for modeling, Working Paper.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2012, Hedge fund stock trading in the financial crisis of 2007–2009, *The Review of Financial Studies* 25, 1–54.

Bernanke, Ben S., and Mark Gertler, 1989, Agency costs, net worth, and business fluctuations, *The American Economic Review* 79, 14–31.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, 1999, The financial accelerator in a quantitative business cycle framework, *Handbook of Macroeconomics* 1, 1341–1393.

Bhamra, Harjoat S., and Raman Uppal, 2009, The effect of introducing a non-redundant derivative on the volatility of stock-market returns when agents differ in risk aversion, *The Review of Financial Studies* 22, 2303–2330.

Bhamra, Harjoat S., and Raman Uppal, 2014, Asset prices with heterogeneity in preferences and beliefs, *The Review of Financial Studies* 27, 519–580.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *The Review of Financial Studies* 22, 4463–4492.

Brunnermeier, Markus K., Thomas M. Eisenbach, and Yuliy Sannikov, 2012, Macroeconomics with financial frictions: A survey Working Paper 18102, National Bureau of Economic Research.

Brunnermeier, Markus K., and Lasse Heje Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201–2238.

Brunnermeier, Markus K., and Yuliy Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104, 379–421.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Chabakauri, Georgy, 2013, Dynamic equilibrium with two stocks, heterogeneous investors, and portfolio constraints, *The Review of Financial Studies* 26, 3104–3141.

Cochrane, John H., 2005, *Asset pricing* (Princeton university press Princeton, NJ).

Coimbra, Nuno, and Hélène Rey, 2017, Financial cycles with heterogeneous intermediaries Working Paper 23245, National Bureau of Economic Research.

Danielsson, Jon, Hyun Song Shin, and Jean-Pierre Zigrand, 2012, Procyclical leverage and endogenous risk Working Paper.

Di Tella, Sebastian, 2017, Uncertainty shocks and balance sheet recessions, *Journal of Political Economy* 125, 2038–2081.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2018, A model of monetary policy and risk premia, *The Journal of Finance* 73, 317–373.

Duffie, Darrel, and Pierre-Louis Lions, 1992, Pde solutions of stochastic differential utility, *Journal of Mathematical Economics* 21, 577 – 606.

Duffie, Darrell, and Larry G. Epstein, 1992, Stochastic differential utility, *Econometrica* 60, 353–394.

Dumas, Bernard, 1989, Two-person dynamic equilibrium in the capital market, *Review of Financial Studies* 2, 157–188.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Gârleanu, Nicolae, and Stavros Panageas, 2015, Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing, *Journal of Political Economy* 123, 670–685.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2011, Margin-based asset pricing and deviations from the law of one price, *The Review of Financial Studies* 24, 1980–2022.

Gatev, Evan, and Philip E. Strahan, 2006, Banks' advantage in hedging liquidity risk: Theory and evidence from the commercial paper market, *The Journal of Finance* 61, 867–892.

Gertler, Mark, and Nobuhiro Kiyotaki, 2010, Chapter 11 - Financial intermediation and credit policy in business cycle analysis, volume 3 of *Handbook of Monetary Economics*, 547 – 599 (Elsevier).

Gertler, Mark, Nobuhiro Kiyotaki, and Andrea Prestipino, 2016, Chapter 16 - Wholesale banking and bank runs in macroeconomic modeling of financial crises, volume 2 of *Handbook of Macroeconomics*, 1345 – 1425 (Elsevier).

Giglio, Stefano, Bryan Kelly, and Seth Pruitt, 2016, Systemic risk and the macroeconomy: An empirical evaluation, *Journal of Financial Economics* 119, 457 – 471.

Gorton, Gary, and Andrew Metrick, 2012, Securitized banking and the run on repo, *Journal of Financial Economics* 104, 425 – 451.

Goyal, Amit, and Ivo Welch, 2008, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *The Review of Financial Studies* 21, 1455–1508.

Gupta, Arun, 2018, The internal capital markets of global dealer banks, Working Paper.

Haddad, Valentin, and Tyler Muir, 2018, Do intermediaries matter for aggregate asset prices? Working Paper.

Hansen, Lars Peter, John C. Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political Economy* 116, 260–302.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1 – 35.

He, Zhiguo, In Gu Khang, and Arvind Krishnamurthy, 2010, Balance sheet adjustments during the 2008 crisis, *IMF Economic Review* 58, 118–156.

He, Zhiguo, and Arvind Krishnamurthy, 2012, A model of capital and crises, *The Review of Economic Studies* 79, 735–777.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103, 732–70.

He, Zhiguo, and Arvind Krishnamurthy, 2019, A macroeconomic framework for quantifying systemic risk, *American Economic Journal: Macroeconomics* , Forthcoming.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357–386.

Judd, Kenneth L., 1992, Projection methods for solving aggregate growth models, *Journal of Economic Theory* 58, 410 – 452.

Judd, Kenneth L., 1998, *Numerical Methods in Economics* (MIT Press).

Kiyotaki, Nobuhiro, and John Moore, 1997, Credit cycles, *Journal of Political Economy* 105, 211–248.

Kogan, Leonid, Igor Makarov, and Raman Uppal, 2007, The equity risk premium and the riskfree rate in an economy with borrowing constraints, *Mathematics and Financial Economics* 1, 1–19.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *The Journal of Finance* 56, 815–849.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175 – 194.

Longstaff, Francis A., and Jiang Wang, 2012, Asset pricing and the credit market, *Review of Financial Studies* 25, 3169–3215.

Ma, Sai, 2017, Heterogeneous intermediaries and asset prices Working Paper.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Moreira, Alan, and Alexi Savov, 2017, The macroeconomics of shadow banking, *The Journal of Finance* 72, 2381–2432.

Nozawa, Yoshio, 2017, What drives the cross-section of credit spreads?: A variance decomposition approach, *The Journal of Finance* 72, 2045–2072.

Protter, Philip E., 2004, *Stochastic Integration and Differential Equations*, second edition (Springer-Verlag).

Rytchkov, Oleg, 2014, Asset pricing with dynamic margin constraints, *The Journal of Finance* 69, 405–452.

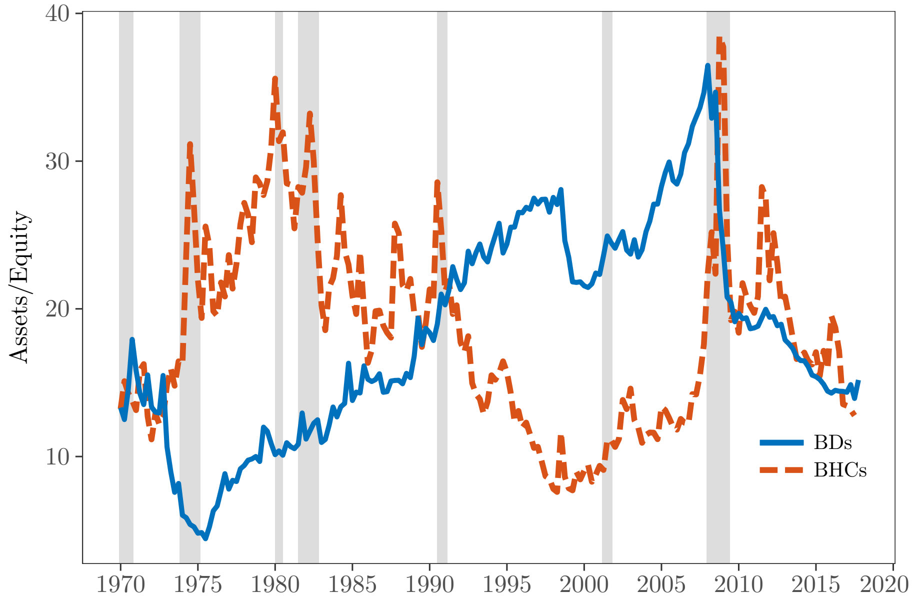
Santos, Tano, and Pietro Veronesi, 2016, Habits and leverage Working Paper 22905, National Bureau of Economic Research.

Schwert, G. William, 1989, Why does stock market volatility change over time?, *The Journal of Finance* 44, 1115–1153.

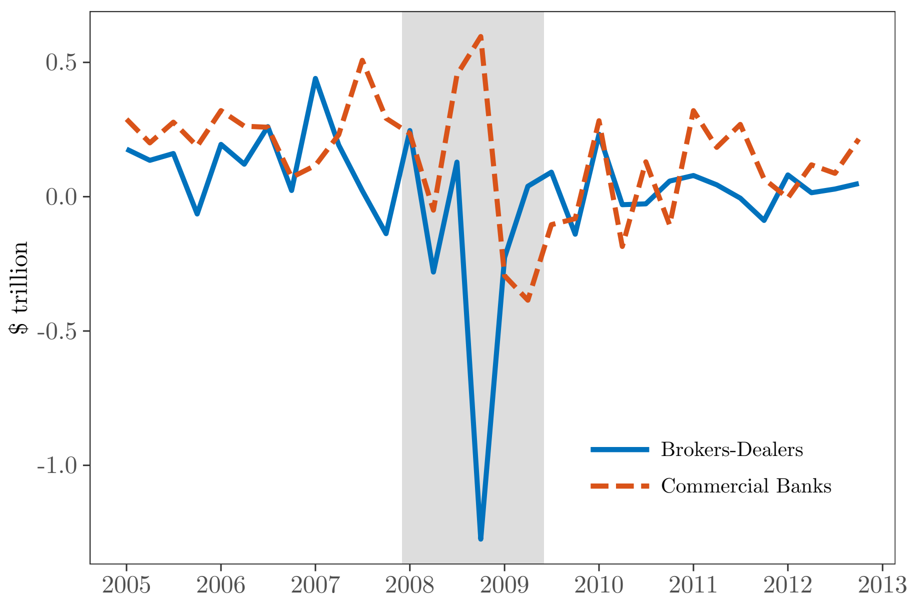
Shanken, Jay, 1992, On the estimation of beta-pricing models, *The Review of Financial Studies* 5, 1–33.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375 – 421.

Wang, Jiang, 1996, The term structure of interest rates in a pure exchange economy with heterogeneous investors, *Journal of Financial Economics* 41, 75 – 110.



(a) Leverage of different financial intermediaries.



(b) Change in total financial assets.

Figure 1. Panel (a) presents time-series of leverage for different financial intermediaries: security broker-dealers (BDs) and bank holding companies (BHCs). Leverage for broker-dealers (solid blue line) is defined as the ratio total financial assets to total equity (total financial assets minus total liabilities) from Table L.130 of the Flow of Funds. BHC leverage (dashed red line) is defined as the ratio of total market assets (book debt plus market equity) to total market equity constructed for publicly-traded holding companies of the NY Fed's primary dealer counterparties using CRSP/Compustat and Datastream, where market equity is outstanding shares times stock price and book debt is total assets minus common equity. Panel (b) presents quarterly change in total financial assets for BDs and Private Depository Institutions (DIs). BDs' (solid blue line) and DIs' (dashed red line) total financial assets are from Tables L.130 and L.110 of the Financial Accounts of the United States (Flow of Funds), respectively. Data is quarterly from 1970Q1 to 2017Q4. The vertical shaded bars indicate NBER recessions.

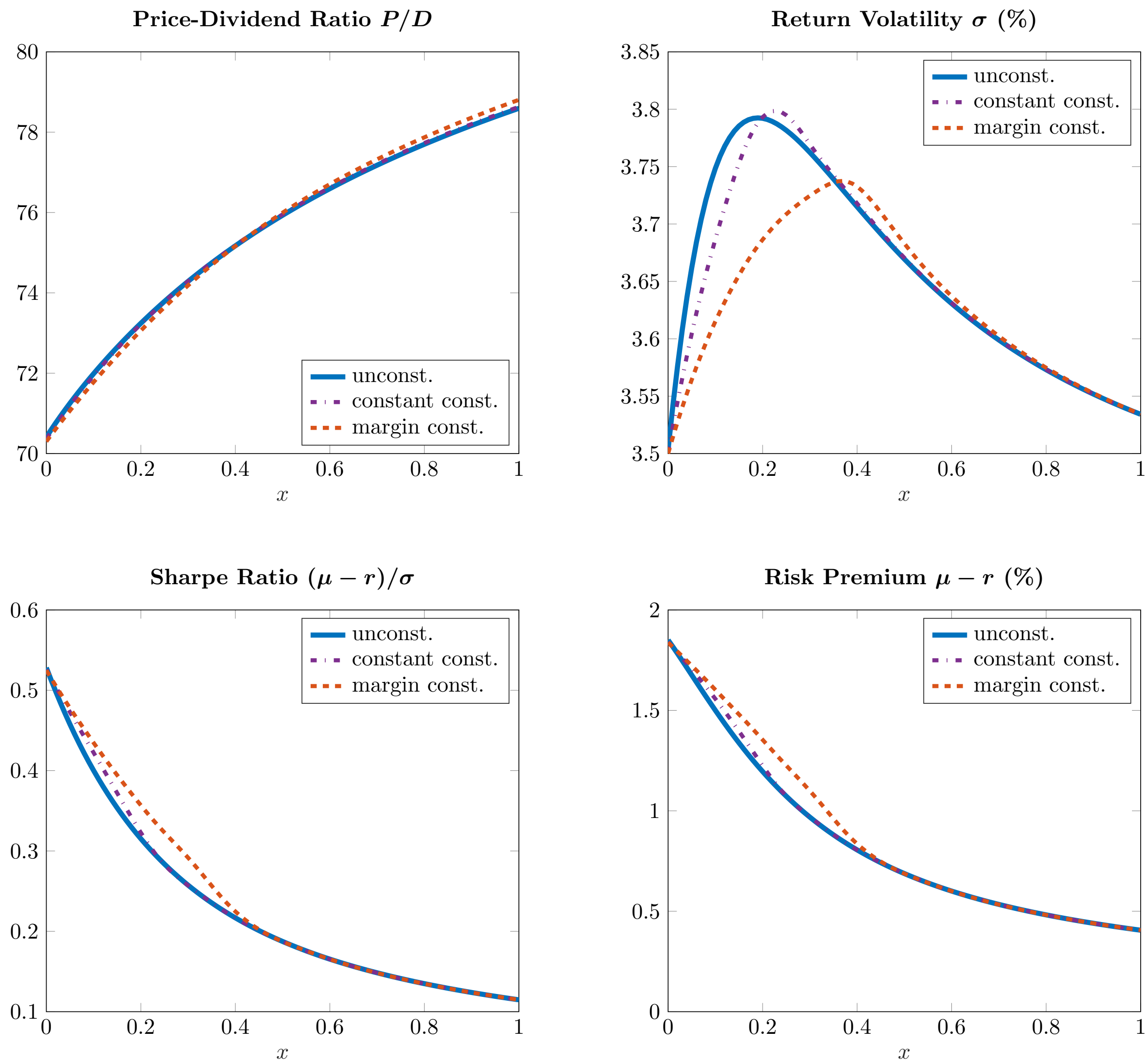


Figure 2. Risk premia, the price of risk, valuation, and volatility. This figure presents price-dividend ratio $1/F$, return volatility σ, Sharpe ratio and risk premium on the endowment claim in constrained and unconstrained equilibria as functions of state variable x (wealth share of the financial sector i.e. type A and B agents) under the benchmark parameters in Table 1. Each quantity is plotted against state variable x_t while the value of the second state variable y_t (wealth share of type A investors, i.e. broker-dealers, in the financial sector) is fixed at 0.56 (its value at the stochastic steady state). The solid blue line corresponds to the unconstrained economy, the dash-dotted purple line corresponds to the economy with a constant portfolio constraint ($\bar{\theta}_t = \bar{m}$), and the dashed red line corresponds to the the economy with a Value-at-Risk (VaR)-type margin constraint ($\bar{\theta}_t = \frac{1}{\alpha\sigma_t}$). Three-dimensional plots are provided in Appendix C.2.

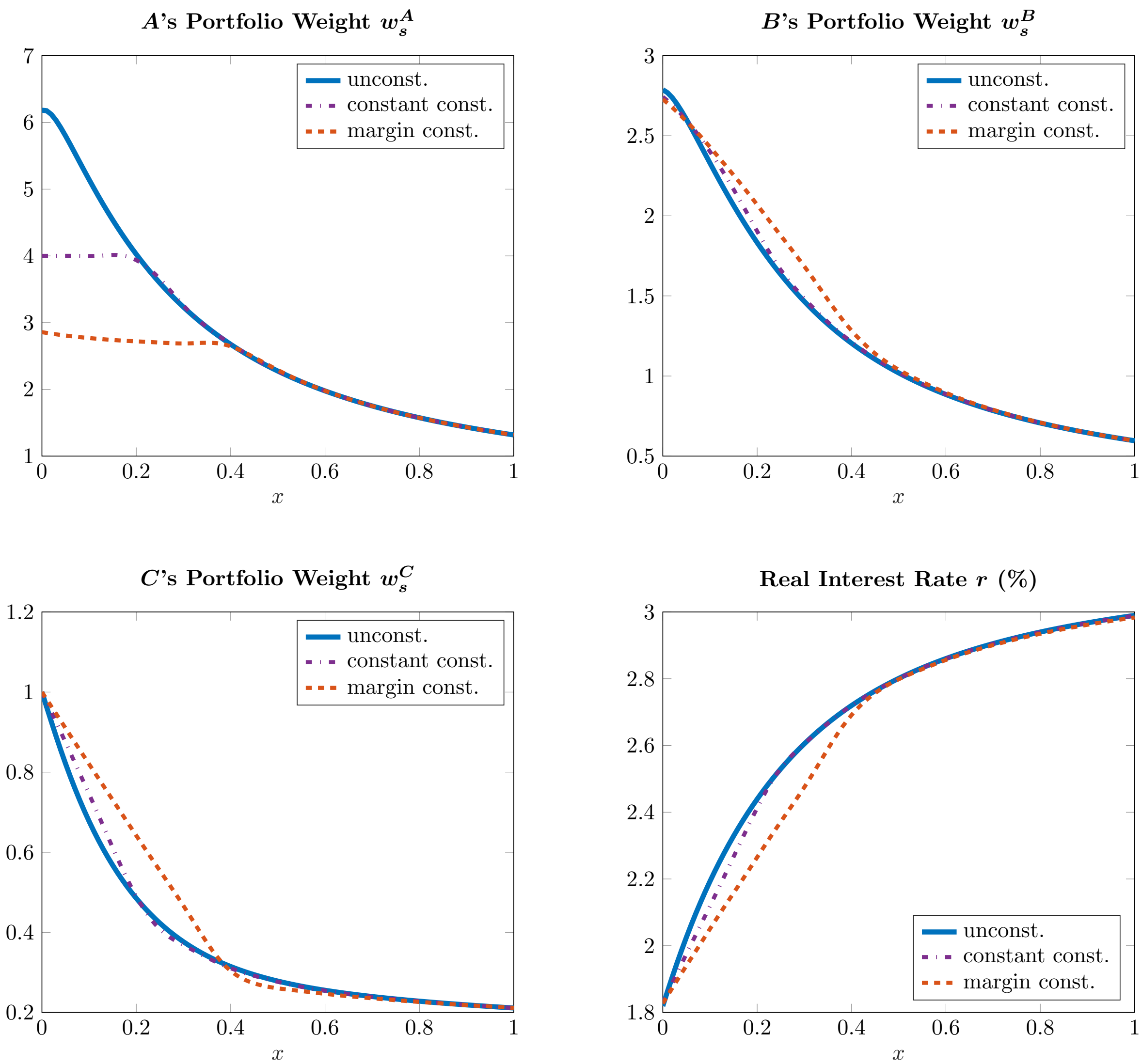


Figure 3. Optimal portfolios and the risk-free rate. This figure presents portfolio weights of each type of agent w_s^A, w_s^B, and w_s^C as well as the real interest rate r_t in constrained and unconstrained equilibria as functions of state variable x (wealth share of the financial sector i.e. A and B agents) under the benchmark parameters in Table 1. Each quantity is plotted against state variable x_t while the value of the second state variable y_t (wealth share of type A investors in the financial sector) is fixed at 0.56 (its value at the stochastic steady state). The solid blue line corresponds to the unconstrained economy, the dash-dotted purple line corresponds to the economy with a constant portfolio constraint ($\bar{\theta}_t = \bar{m}$), and the dashed red line corresponds to the the economy with a Value-at-Risk (VaR)-type margin constraint ($\bar{\theta}_t = \frac{1}{\alpha \sigma_t}$). Three-dimensional plots are provided in Appendix C.2.

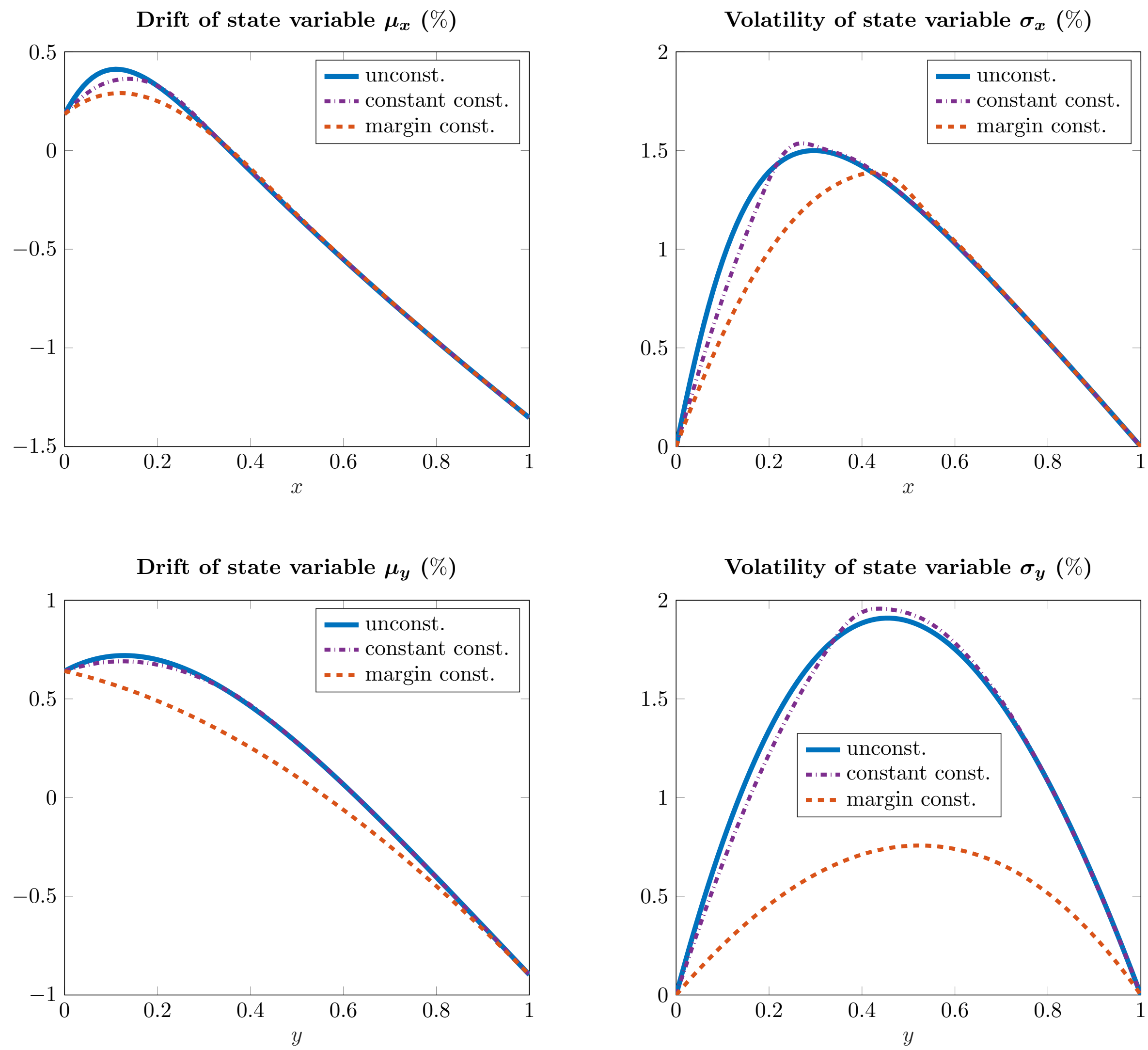


Figure 4. Dynamics of the endogenous state variables. This figure presents dynamics of the state variables x and y (wealth share of the financial sector i.e. A and B agents, and wealth share of type A investors in the financial sector, respectively) in constrained and unconstrained equilibria under the benchmark parameters in Table 1. Drift and volatility of state variable x (i.e. μ_x, σ_x) are plotted as functions of x while the value of the state variable y is fixed at 0.56. Drift and volatility of state variable y (i.e. μ_y, σ_y) are plotted as functions of y while the value of the state variable x is fixed at 0.25. The solid blue line corresponds to the unconstrained economy, the dash-dotted purple line corresponds to the economy with a constant portfolio constraint ($\bar{\theta}_t = \bar{m}$), and the dashed red line corresponds to the the economy with a Value-at-Risk (VaR)-type margin constraint ($\bar{\theta}_t = \frac{1}{\alpha \sigma_t}$). Three-dimensional plots are provided in Appendix C.2.

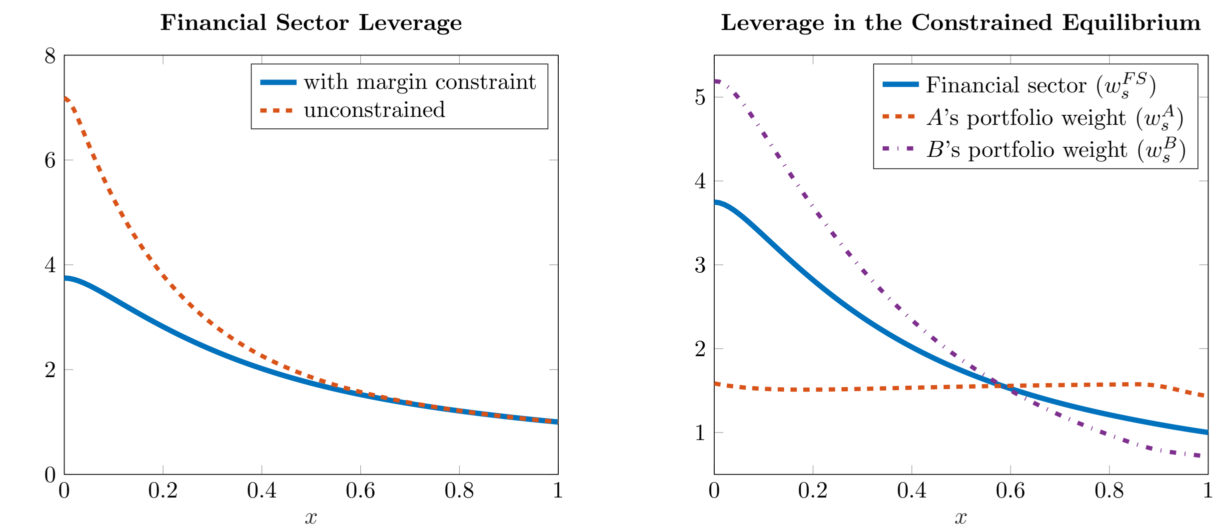


Figure 5. Cyclical properties of intermediary leverage. This figure presents optimal intermediary leverage in the unconstrained and constrained equilibria under parameters listed in Table 1. The left panel plots leverage of the financial sector (equation 28) in the unconstrained equilibrium (dashed red line) and the model with time-varying margin constraints, $\bar{\theta}_t = \frac{1}{\alpha\sigma_t}$ (solid blue line). The right panel plots intermediary leverage in the main model with endogenous margin constraints. The solid blue line corresponds to leverage of the financial sector (w_s^{FS}), the dashed red line presents broker-dealers' leverage (w_s^A), and the dash-dotted purple line corresponds to leverage of bank holding companies (w_s^B). Each quantity is plotted against state variable x_t (wealth share of the financial sector i.e. A and B agents) while the value of the second state variable y_t (wealth share of type A investors in the financial sector) is held fixed at 0.56 (its stochastic steady state value).

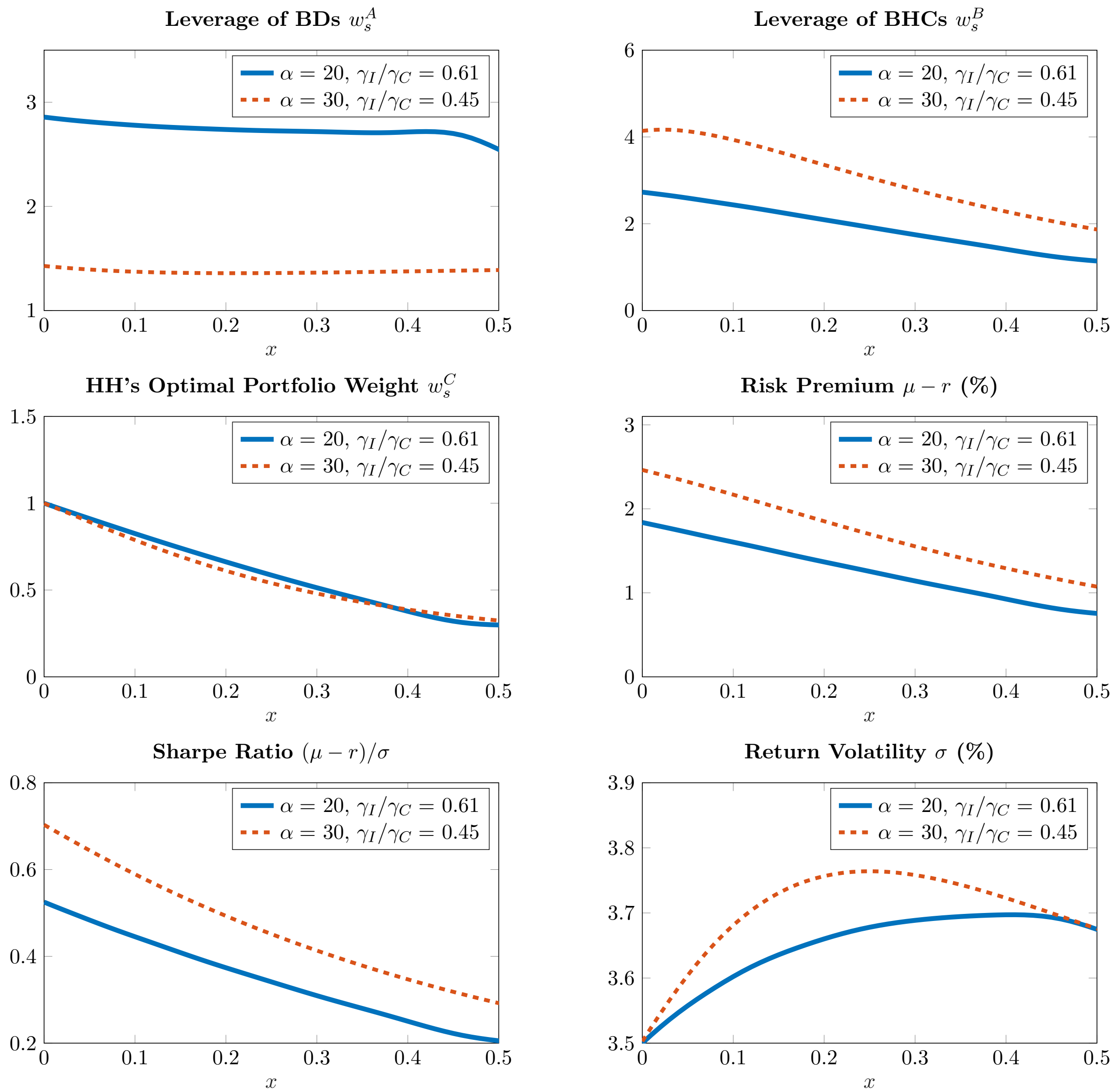


Figure 6. Asset reallocation within the financial sector. This figure presents portfolio weights for dealers, holding companies, and households (A, B, and C types, respectively), as well as, the risk premium and Sharpe ratio of the risky claim on the aggregate endowment, and the volatility of the risky asset return in the baseline model (solid blue line) and a model with tighter margin constraints and less risk averse financial sector (dashed red line). The changes in tightness of the margin constraint (parameter α) and relative risk aversion of the financial and household sectors (γ_I/γ_C) are such that leverage of A (B) types is reduced (increased) by approximately 47% (72%): changes documented during the Great Recession in Figure 1a. Each quantity is plotted against state variable x (wealth share of the financial sector i.e. A and B agents) while value of the state variable y (wealth share of dealers i.e. type A investors in the financial sector) is fixed at 0.56, its value at the stochastic steady state. Parameters for the baseline model are presented in Table 1.

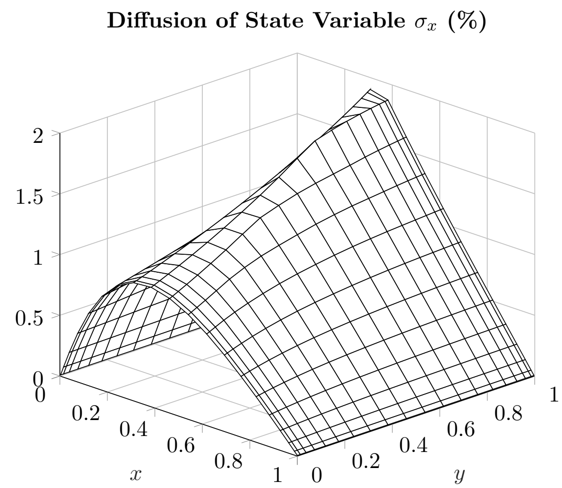


Figure 7. State variable diffusions. This figure presents the diffusions of state variables x and y (σ_x and σ_y, respectively) in the economy with time-varying margin constraints as functions of state variable x (wealth share of the financial sector i.e. type A and B agents) and y_t (wealth share of type A agents in the financial sector) under the benchmark parameters in Table 1.

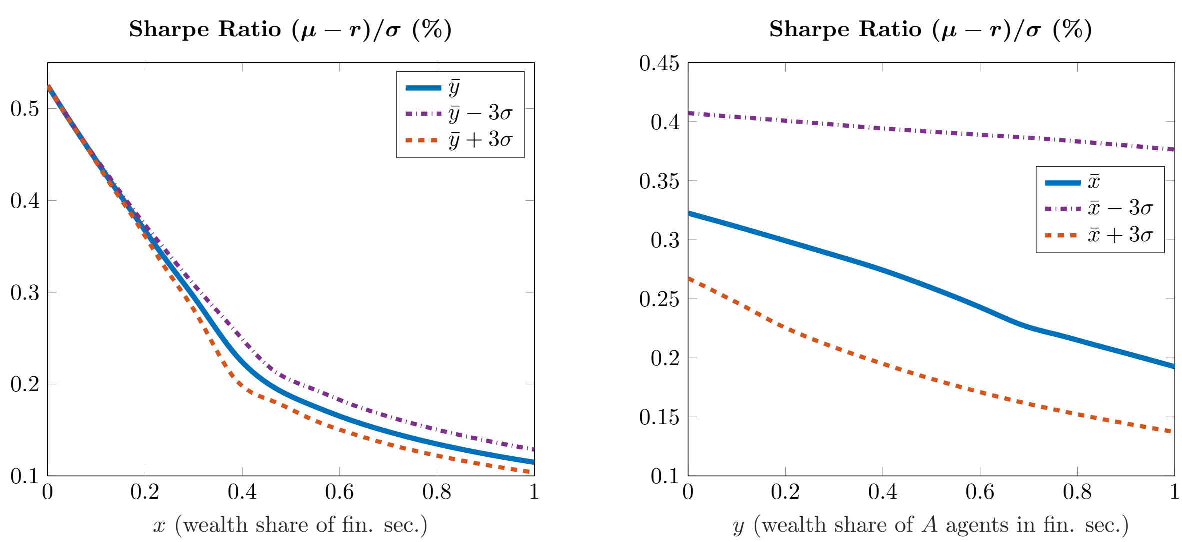


Figure 8. Price of risk. This figure presents the Sharpe ratio of the risk asset in the economy with time-varying margin constraints as functions of state variable x (wealth share of the financial sector i.e. type A and B agents) and y_t (wealth share of type A agents in the financial sector) under the benchmark parameters in Table 1.

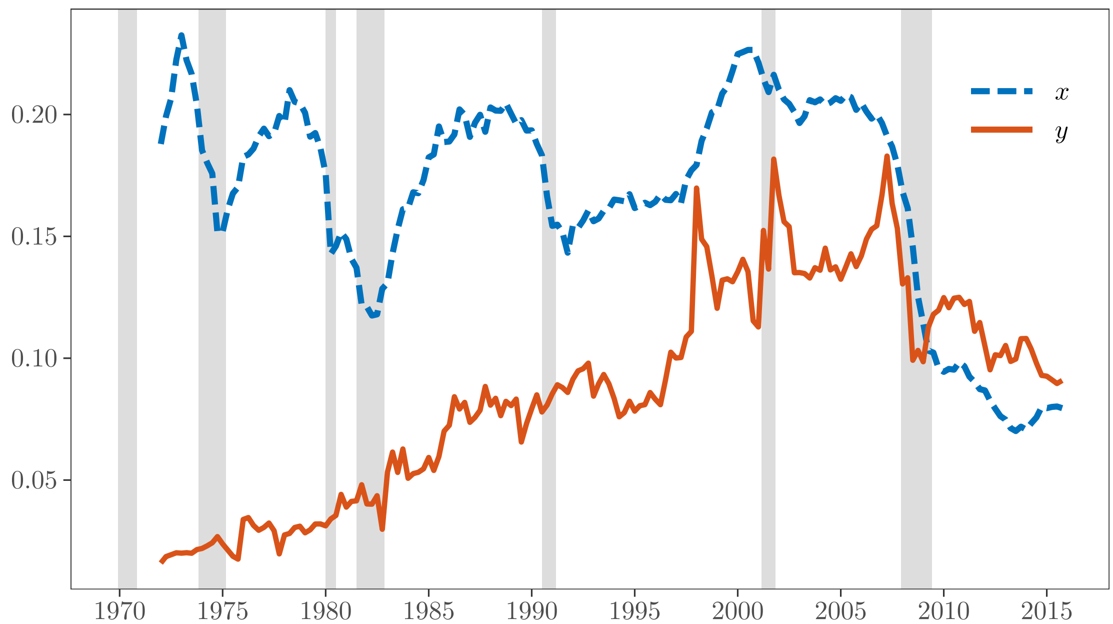


Figure 9. State variables x and y in the data. This figure presents the three-month moving averages of monthly wealth share of the financial sector, x^{data}, and quarterly equity share of the broker-dealers in the financial sector, y^{data}, defined in equations (29) and (30), respectively. Financial sector is identified as firms in the CRSP universe for which the first two digits of the header SIC code (HSICCD in CRSP) equals 60–67. Book equity for BDs and depository institutions are computed from the Flow of Funds Tables L.130 and L.110, respectively. Sample period is from 1970 to 2018. Both time-series are standardized to have zero mean and unit standard deviation for illustration. The vertical shaded bars indicate NBER recessions.

Table 1. Parameter values for the endowment economy model.
This tables reports parameter values used in calibrating the model. The model is set in continuous time. So, the values correspond to annual values rather than the typical quarterly ones seen in discrete time calibrations.

	Parameter	Choice	Target
	Preferences		
ψ_A	EIS of type A	1.5	Literature
ψ_B	EIS of type B	1.5	Literature
ψ_C	EIS of type C	1.5	Literature
γ_A	Risk aversion of type A	2.5	Free parameter
γ_B	Risk aversion of type B	5.5	Dealer/bank leverage ratio
γ_C	Risk aversion of type C	15	Mean Sharpe ratio
ρ	Rate of time preference	0.001	Literature
	Endowment and Demography		
μ_D	Endowment growth rate	0.022	US consumption data
σ_D	Endowment volatility	0.035	US consumption data
κ	Agents birth/death rate	0.015	Literature
$\bar{u}$	Population share of type A	0.05	Intermediary and BD wealth shares
$\bar{v}$	Population share of type B	0.07	Intermediary and BD wealth shares
	Margin Constraint		
$\bar{m}$	Constant leverage constraint	4	Literature
α	Tightness of the dynamic constraint	10	30-day 99% VaR

Table 2. State variables statistics.
This table reports statistics for empirical proxies for model's two state variables in level and changes (Innov.). AC(j) represent j^{th} autocorrelation. Data is quarterly from 1970Q1 to 2017Q4.

	x^{data}		y^{data}	
	Level	Innov.	Level	Innov.
Mean	0.141	0.000	0.191	0.000
Std Dev	0.037	0.007	0.068	0.018
AC(1)	0.966	0.003	0.958	−0.185
AC(2)	0.933	0.046	0.930	0.087
AC(3)	0.899	0.105	0.896	−0.087
AC(4)	0.861	0.004	0.870	−0.075
AC(5)	0.825	−0.025	0.851	0.071

Table 3. Predictive regressions: $\boldsymbol{y_t}$.
This table provides results for one-year ahead predictive regressions according to $R^i_{t+1\to t+4} - r^f_t = \gamma_0 + \gamma^i_y\, y_t + \varepsilon^i_{t+1\to t+4}$, using lagged equity share of broker-dealers in the financial sector, the empirical proxy for state variable y defined in equation (30) which captures the composition of the financial sector, as the predictor of interest. The dependent variables are excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}), mean excess return on 25 Fama-French size and book-to-market (FF25_{t+1}), 10 momentum (Mom_{t+1}) portfolios, 10 maturity-sorted US government and 10 US corporate bond portfolios sorted on yield spreads (US bonds_{t+1}), mean excess returns on six sovereign bonds (Sov. bonds_{t+1}), 54 portfolios of S&P 500 index options sorted on moneyness and maturity (Options_{t+1}), 20 CDS portfolios sorted by spreads (CDS_{t+1}), 23 commodity (Commod._{t+1}), and 12 foreign exchange (FX_{t+1}) portfolios. Size/book-to-market and momentum portfolios and the risk-free rate data are from Ken French's website. Data on sovereign bonds, options, CDS, commodities, and FX portfolios are from He et al. (2017). The sample quarterly from 1974Q2 to 2017Q2 for market, FF25 and momentum portfolios, and to 2012Q4 for HKM assets. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable:								
	Mkt	FF25	Mom	US bonds	Sov. bonds	Options	CDS	Commod.	FX
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
y_t^{data}	-1.16^{***}	-0.98^{***}	-0.77^{**}	-0.11	-1.09^{**}	-1.57^{**}	-0.04	-0.40	0.37
	(0.38)	(0.28)	(0.33)	(0.14)	(0.46)	(0.61)	(0.13)	(0.45)	(0.31)
Const	0.39^{***}	0.34^{***}	0.27^{***}	0.09^{**}	0.39^{***}	0.46^{***}	0.03	0.14	-0.08
	(0.08)	(0.05)	(0.06)	(0.04)	(0.11)	(0.14)	(0.03)	(0.11)	(0.08)
Obs	173	173	173	152	62	100	44	102	132
R^2	0.10	0.11	0.08	0.01	0.15	0.12	0.005	0.01	0.05
Adjusted R^2	0.09	0.11	0.07	0.01	0.14	0.11	-0.02	0.003	0.05

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4. Predictive regressions: y_t and Controls.
This table provides results for one-year ahead predictive regressions according to $R^i_{t+1\to t+4} - r^f_t = \gamma_0 + \gamma^i_y\, y_t + \gamma^i_{\text{Ctrl}}\,\text{Ctrl}_t + \varepsilon^i_{t+1\to t+4}$, using lagged equity share of broker-dealers in the financial sector, the empirical proxy for state variable y defined in equation (30) which captures the composition of the financial sector, as the predictor of interest. Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: wealth share of the aggregate financial sector (x from the model defined in 29), *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variables are excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}), mean excess return on 25 Fama-French size and book-to-market (FF25_{t+1}), 10 momentum (Mom_{t+1}) portfolios, 10 maturity-sorted US government and 10 US corporate bond portfolios sorted on yield spreads (US bonds_{t+1}), mean excess returns on six sovereign bonds (Sov. bonds_{t+1}), 54 portfolios of S&P 500 index options sorted on moneyness and maturity (Options_{t+1}), 20 CDS portfolios sorted by spreads (CDS_{t+1}), 23 commodity (Commod._{t+1}), and 12 foreign exchange (FX_{t+1}) portfolios. Size/book-to-market and momentum portfolios and the risk-free rate data are from Ken French's website. Data on sovereign bonds, options, CDS, commodities, and FX portfolios are from He et al. (2017). The sample quarterly from 1990Q1 to 2017Q3 for market, FF25 and momentum portfolios, and to 2012Q4 for HKM assets. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable:									
	Mkt		FF25	FFmom	US bonds	Sov. bonds	Options	CDS	Commod.	FX
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
y_t		-1.77^{***}	-1.72^{***}	-1.80^{***}	0.43^{***}	-0.27	-2.14^{***}	-0.18	-1.15	0.44
		(0.46)	(0.51)	(0.45)	(0.18)	(0.64)	(0.74)	(0.16)	(0.74)	(0.40)
Const	0.04	-0.01	0.23	0.14	0.14^{***}	0.54^{***}	0.03	0.001	-0.06	-0.09
	(0.19)	(0.13)	(0.15)	(0.14)	(0.03)	(0.11)	(0.10)	(0.03)	(0.13)	(0.07)
Ctrls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs	109	109	109	109	88	62	84	44	88	76
R^2	0.31	0.46	0.31	0.44	0.35	0.26	0.57	0.38	0.16	0.41
Adj R^2	0.28	0.43	0.27	0.40	0.30	0.18	0.54	0.28	0.10	0.36

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5. Cross-sectional asset pricing tests.
This table presents pricing results for the 25 size/book-to-market, 10 momentum, 10 maturity-sorted Treasury bond portfolios from CRSP with maturities in six month intervals up to five years, and 10 US corporate bond portfolios sorted on yield spreads from Nozawa (2017). The table reports the prices of risk and test diagnostics, including mean absolute pricing errors (MAPEs), and adjusted R^2s, and a χ^2 statistic and p-value that tests whether the pricing errors are jointly zero. Shanken (1992)-corrected and Fama and MacBeth (1973) t-statistics (t-Shanken and t-FM, respectively) are reported in parentheses. Heterogeneous intermediary factor (HIFac) is defined as the AR(1) innovations in the wealth share of dealers, scaled by their lagged wealth share according to equation (32). The AEM leverage factor (AEMLevFac) is defined as the seasonally-adjusted growth rate in broker-dealer leverage from Table L.130 of the Flow of Funds. HKM capital factor (HKMFac) is the shock to intermediary capital ratio in He et al. (2017), defined as the ratio of total market equity to total market assets (book debt plus market equity) for bank holding companies of New York Fed's primary dealer counterparties. MktRF is the excess return on CRSP value-weighted portfolio from Ken French's website. The sample is quarterly from 1970Q1 to 2017Q4. Returns and risk premia are reported in percentage per year (quarterly percentages multiplied by four).

	(1)	(2)	(3)	(4)	(5)	(6)
HIFac	38.27*	57.35**		34.87*		52.47*
t-Shanken	(1.83)	(2.11)		(1.78)		(1.74)
t-FM	(2.24)	(3.11)		(2.02)		(2.82)
MktRF		3.86			6.93*	4.61
t-Shanken		(1.20)			(1.97)	(1.38)
t-FM		(1.27)			(2.20)	(1.52)
AEMLevFac			32.50***	21.66**		
t-Shanken			(2.68)	(2.42)		
t-FM			(3.73)	(3.14)		
HKMFac					12.55**	11.80**
t-Shanken					(2.57)	(2.12)
t-FM					(3.96)	(3.68)
Intercept	3.47***	4.14***	5.32**	3.01**	2.77*	3.99***
t-Shanken	(2.92)	(2.90)	(2.03)	(2.32)	(1.84)	(2.38)
t-FM	(3.62)	(4.32)	(2.68)	(3.11)	(2.90)	(4.19)
Observations	55	55	55	55	55	55
Adjusted R^2	0.61	0.61	0.39	0.72	0.63	0.69
MAPE, %	1.83	1.84	2.96	1.58	1.89	1.67
$\chi^2(N-K)$	195.39	133.53	151.78	167.17	121.35	95.34
p-value	0.00	0.00	0.00	0.00	0.00	0.00
Note:				*p<0.1; **p<0.05; ***p<0.01		

Appendix

A Proof of Propositions

Proof of Proposition 1

Proof. State variable x is the wealth share of financial sector: $x_t = \frac{W^A+W^B}{W}$ and state variable y is wealth share of type A agents in the financial sector: $y_t = \frac{W^A}{W^A+W^B}$, where W, W^A, and W^B are the aggregate wealth, wealth of A, and B agents, respectively.

Dynamics of x_t: From equation (5), W^A has the following law of motion

$$\frac{dW^A}{W^A} = \left(r + w_s^A(\mu - r) - c_A\right) dt + w_s^A \sigma\, dZ,$$

and W^B has the similar law of motion.

The law of motion for the numerator, $W^A + W^B$, will be

$$\begin{aligned}\frac{d\left(W^A + W^B\right)}{W^A + W^B} &= \left[r + \left(yw_s^A + (1-y)w_s^B\right)(\mu - r) - (yc_A + (1-y)c_B)\right] dt \\ &+ \left(yw_s^A + (1-y)w_s^B\right)\sigma dZ_t\end{aligned}$$

Define wealth share of agents A and B as $u \equiv W^A/W = xy$, and $v \equiv W^B/W = x(1-y)$, respectively.[56] Since the aggregate wealth is $W = W^A + W^B + W^C$, the law of motion for the denominator is

$$\begin{aligned}\frac{dW}{W} &= \left[r + \underbrace{\left(xyw_s^A + x(1-y)w_s^B + (1-x)w_s^C\right)}_{=1 \text{ (by stock market-clearing)}}(\mu - r) - \underbrace{\left(xyc_A + x(1-y)c_B + (1-x)c_C\right)}_{=F \text{ (by goods market-clearing)}}\right] dt \\ &+ \left[\underbrace{xyw_s^A + x(1-y)w_s^B + (1-x)w_s^C}_{=1}\right]\sigma\, dZ_t \\ &= \left[r + (\mu - r) - F\right] dt + \sigma\, dZ_t\end{aligned}$$

From Ito's lemma for ratio of two stochastic processes,

$$\begin{aligned}\frac{dx}{x} &= \kappa(\bar{x} - x)dt + \left[\left(yw_s^A + (1-y)w_s^B - 1\right)(\mu - r - \sigma^2) - yc_A - (1-y)c_B + F\right] dt \\ &+ \left(yw_s^A + (1-y)w_s^B - 1\right)\sigma\, dZ_t\end{aligned}$$

Thus from the dynamics of x in equation (12) we have

$$\begin{aligned}\mu_x &= \left(yw_s^A + (1-y)w_s^B - 1\right)(\mu - r - \sigma^2) - yc_A - (1-y)c_B + F \\ \sigma_x &= \left(yw_s^A + (1-y)w_s^B - 1\right)\sigma\end{aligned}$$

Dynamics of y_t: The numerator of y is W^A, and its denominator is $\left(W^A + W^B\right)$ which its law

[56] Agent C's wealth share will then be $1 - u - v$.

of motion is calculated above. So from Ito's lemma for a ratio, we get

$$\frac{dy}{y} = \kappa(\bar{y} - y)dt + (1 - y)\left[\left(w_s^A - w_s^B\right)(\mu - r) - c_A + c_B - \left(yw_s^A + (1 - y)w_s^B\right)\left(w_s^A - w_s^B\right)\sigma^2\right] dt$$
$$+ (1 - y)\left(w_s^A - w_s^B\right)\sigma\, dZ_t$$

Thus from the dynamics of y in equation (13) we have

$$\mu_y = \left(w_s^A - w_s^B\right)(\mu - r) - c_A + c_B - \left(yw_s^A + (1 - y)w_s^B\right)\left(w_s^A - w_s^B\right)\sigma^2$$
$$\sigma_y = \left(w_s^A - w_s^B\right)\sigma$$

□

Proof of Proposition 2

Proof. Since σ_x and σ_y are finite, we trivially get

$$\lim_{x\to 0} x\,\sigma_x = 0,\ \forall y \quad \text{and} \quad \lim_{y\to 0} y(1-y)\,\sigma_y = \lim_{y\to 1} y(1-y)\,\sigma_y = 0,\ \forall x.$$

We only need to show $\lim_{x\to 1} x\,\sigma_x = 0\ \forall y$. The market clearing condition for the risky asset when $x \to 1$ becomes $yw_s^A + (1-y)w_s^B = 1$.

So, from the expression for σ_x in equation (16), we have:

$$x\,\sigma_x = x\left[yw_s^A + (1-y)w_s^B - 1\right]\sigma$$

which goes to zero as $x \to 1$ for all y from the stock market-clearing. □

Proof of Proposition 3

Proof. We can write agent i's optimization problem in equation (8) as

$$0 = \max_{c_i, w_s^i}\left\{f_i\left(c_{i,t}, V_{i,t}\right)dt + \mathbb{E}_t\left[dV_{i,t}\right]\right\}$$

Using Ito's lemma we have

$$\mathbb{E}_t[dV_i] = V_{i,W_i}\mathbb{E}_t[dW_i] + \frac{1}{2}V_{i,W_iW_i}\mathbb{E}_t\left[dW_i^2\right] + V_{i,J_i}\mathbb{E}_t[dJ_i] + \frac{1}{2}V_{i,J_iJ_i}\mathbb{E}_t\left[dJ_i^2\right] + V_{i,W_iJ_i}\mathbb{E}_t[dW_i dJ_i]$$

where V_{i,W_i} and V_{i,W_iW_i} are the first and second partial derivatives of V_i with respect to W_i (similarly for V_{i,J_i}, V_{i,J_iJ_i}, and V_{i,W_iJ_i}.) Also posit the following Ito process for marginal value of wealth J_i:

$$\frac{dJ_i}{J_i} = \mu_{J_{i,t}}dt + \sigma_{J_{i,t}}dZ,$$

with adapted processes $\mu_{J_{i,t}} = \mu_{J_i}(x_t, y_t)$ and $\sigma_{J_{i,t}} = \sigma_{J_i}(x_t, y_t)$. I will drop t subscripts for notational simplicity.

Using Ito's lemma, we can find the drift and diffusions μ_{J_i} and σ_{J_i}

$$\begin{aligned}\mu_{J_i} &= \frac{J_{i,x}}{J_i}\left[\kappa(\bar{x}-x)+x\,\mu_x\right]+\frac{J_{i,y}}{J_i}\left[\kappa(\bar{y}-y)+y(1-y)\mu_y\right]\\&\quad+\frac{1}{2}\frac{J_{i,xx}}{J_i}x^2\sigma_x^2+\frac{J_{i,xy}}{J_i}xy(1-y)\sigma_x\sigma_y+\frac{1}{2}\frac{J_{i,yy}}{J_i}y^2(1-y)^2\sigma_y^2 \qquad (A.1)\\ \sigma_{J_i} &= \frac{J_{i,x}}{J_i}x\,\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\sigma_y \qquad (A.2)\end{aligned}$$

Plugging in the felicity function $f(C,U)$ in (3) and the conjecture for value function V_i in (21) into the HJB equation above, using the budget constraint in (5) and the law of motion for J_i in (A.1) and (A.2), and dropping $W_i^{1-\gamma_i}J_i^{\frac{1-\gamma_i}{1-\psi_i}}$ and dt terms yields

$$\begin{aligned}0 = \max_{c_i,w_s^i}\ &\frac{1}{1-1/\psi_i}\left[\left(\frac{c_i}{J_i^{1/(1-\psi_i)}}\right)^{1-1/\psi_i}-(\rho+\kappa)\right]+\left[r-c_i+\kappa+w_s^i(\mu-r)-\frac{\gamma_i}{2}\left(w_s^i\right)^2\sigma^2\right]\\&+\left(\frac{1}{1-\psi_i}\right)\left\{\frac{J_{i,x}}{J_i}\left[\kappa(\bar{x}-x)+x\mu_x\right]+\frac{J_{i,y}}{J_i}\left[\kappa(\bar{y}-y)+y(1-y)\mu_y\right]\right.\\&\left.+(1-\gamma_i)\left(\frac{J_{i,x}}{J_i}x\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\sigma_y\right)w_s^i\sigma\right\}+\frac{1}{2}\left(\frac{1}{1-\psi_i}\right)\left[\left(\frac{\psi_i-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\mu_y\right)^2\right.\\&\left.+\frac{J_{i,xx}}{J_i}x^2\sigma_x^2+2\frac{J_{i,xy}}{J_i}xy(1-y)\sigma_x\sigma_y+\frac{J_{i,yy}}{J_i}y^2(1-y)^2\sigma_y^2\right]+\lambda_i\left(\bar{\theta}_t-w_s^i\right),\end{aligned}$$

where λ_i is proportional to the Lagrange multiplier on the time-varying margin constraint. The first-order condition for consumption-wealth ratio and portfolio share will lead to equations (23) and (25):

$$c_i = J_i \qquad (A.3)$$

$$w_s^i = \frac{1}{\gamma_i}\left[\frac{\mu-r}{\sigma^2}+\left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\,\frac{\sigma_x}{\sigma}+\frac{J_{i,y}}{J_i}y(1-y)\frac{\sigma_y}{\sigma}\right)\right]-\frac{1}{\gamma_i\sigma^2}\lambda_i \qquad (A.4)$$

When the margin constraint for agent i is slack, $\lambda_i=0$ and we have

$$w_s^{i,*} = \frac{\mu-r}{\gamma_i\sigma^2}+\frac{1}{\gamma_i}\left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\,\frac{\sigma_x}{\sigma}+\frac{J_{i,y}}{J_i}y(1-y)\frac{\sigma_y}{\sigma}\right)$$

When the margin constraint for agent i is binding, λ_i is strictly positive and $w_s^{i,const}=\bar{\theta}_t$.

Plugging in the $w_s^{i,const}$ into (A.4), we get the expression for the multiplier on the time-varying margin constraint:

$$\lambda_i = (\mu-r)+\left(\frac{1-\gamma_i}{1-\psi_i}\right)\left(\frac{J_{i,x}}{J_i}x\sigma_x+\frac{J_{i,y}}{J_i}y(1-y)\sigma_y\right)\sigma-\gamma_i\sigma^2\bar{\theta}_t \qquad (A.5)$$

□

B Numerical Procedure

The computation of equilibrium is reduced to solving three second-order PDEs for functions J_i for $i \in \{A, B, C\}$.[57] I use Chebyshev orthogonal collocation method to solve the model.[58] The HJB equation for agent i can be written as the following functional equation:

$$\mathcal{H}_i(J_i) = \mathbf{0}.$$

I express marginal value of wealth functions $J_A(x,y)$, $J_B(x,y)$ and $J_C(x,y)$ as bivariate Chebyshev polynomials of order N (I use $N = 20$), that is, I approximate J_i with tensor product of Chebyshev polynomials of order N:

$$\widehat{J}_i(x,y) = \sum_{j=0}^{N}\sum_{k=0}^{N} a^i_{jk}\psi_j(\omega_j(x))\psi_k(\omega_k(y)),\ i \in \{A, B, C\}. \tag{B.1}$$

where ψ_j is the Chebyshev polynomial of degree $j = 0, 1, \ldots, N$, called the *basis function*, $\Psi_{jk}(x,y) = \psi_j(x)\psi_k(y)$ is a tensor product basis, $\{a^i_{jk}\}^N_{j,k=1}$ are unknown coefficients for agent i, and ω_j's are the Chebyshev nodes (collocation points) defined below.

I then plug in $\widehat{J}_i$ into the HJB equation for agent i to form the *residual equation*:

$$\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right) = \mathcal{H}_i\left(\widehat{J}_i\right),$$

and find the vector of coefficients $\boldsymbol{a}^i$ that makes the residual equation as close to $\mathbf{0}$ as possible given some objective function $\rho\left(\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right), \mathbf{0}\right)$:

$$\boldsymbol{a}^i = \arg\min_{\boldsymbol{a}^i}\ \rho\left(\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right), \mathbf{0}\right)$$

The most common objective function is a *weighted residual* given some weight functions $\phi_j : \Omega \to \mathbb{R}^m$:

$$\rho\left(\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right), \mathbf{0}\right) = \begin{cases} 0 & \text{if} \displaystyle\iint_{\Omega\times\Omega} \phi_j(x)\phi_k(y)\mathcal{R}_i\left(\cdot \mid \boldsymbol{a}^i\right) dx\, dy = \mathbf{0},\ \text{for } j,k = 1,\ldots,N \\ 1 & \text{otherwise} \end{cases}$$

In the pseudo-spectral (or collocation) method, the weight functions are chosen as: $\phi_j(x) = \delta(x - x_i)$ where δ is the dirac delta function and x_i's are the collocation points. In the *orthogonal collocation* method, which I use to solve the model, the basis functions are a set of orthogonal Chebyshev polynomials and collocation points are given by the roots of the N^{th} polynomial.

[57] Duffie and Lions (1992) show existence and uniqueness of infinite-horizon stochastic differential utility by partial differential equation techniques in a Markov diffusion setting.

[58] For more details, see Judd (1992, 1998) and Computational Tools & Macroeconomic Applications, NBER Summer Institute 2011 Methods Lectures, Lawrence Christiano and Jesus Fernandez-Villaverde, Organizers.

Chebyshev polynomials of degree n can be easily defined recursively:

$$\begin{aligned}\psi_0(\omega) &= 1 \\ \psi_1(\omega) &= x \\ \psi_{n+1}(\omega) &= 2\omega\psi_n(\omega) - \psi_{n-1}(\omega)\end{aligned} \tag{B.2}$$

As mentioned above, the collocation points are the N zeros of the Chebyshev polynomial of order $N, (\psi_N(\omega_j) = 0)$, and are given by the following expression

$$\omega_j = \cos\left(\frac{2j-1}{2n}\pi\right),\ j = 1, \ldots, N.$$

These roots are clustered quadratically towards ± 1. Chebyshev polynomials are defined on $\omega_i \in [-1, 1]$. Since the state variables $x, y \in [0, 1]$ in my model, I use the linear transformation $x_j = (1 + \omega_j)/2$.[59]

I calculate the derivatives of these functions as well as the state variable dynamics, agents' portfolio choice, risky asset return and volatility using the relevant equilibrium expressions. I then plug these quantities into the HJB equations (22) and project the resulting residuals onto the complete set of Chebyshev polynomials up to order N. I use the built-in MATLAB function `fsolve` to find the coefficients of J_i polynomials that make the projected residuals equal to zero. This results in a highly accurate solution for coefficients in the $\widehat{J_i}$ functions with errors in the order of 10^{-20}.

The numerical algorithm is summarized below.

1. From goods market-clearing conditions and differentiating it with respect to the state variable, we get expressions for dividend yield F and its derivatives with respect to x and y.

$$\begin{aligned}
&F = xyJ_A + x(1-y)J_B + (1-x)J_C, \\
&F_x = yJ_A + (1-y)J_B - J_C + xyJ_{A,x} + x(1-y)J_{B,x} + (1-x)J_{C,x}, \\
&F_y = xJ_A - xJ_B + xyJ_{A,y} + x(1-y)J_{B,y} + (1-x)J_{C,y}, \\
&F_{xx} = 2yJ_{A,x} + 2(1-y)J_{B,x} - 2J_{C,x} + xyJ_{A,xx} + x(1-y)J_{B,xx} + (1-x)J_{C,xx}, \\
&F_{yy} = 2xJ_{A,y} - 2xJ_{B,y} + xyJ_{A,yy} + x(1-y)J_{B,yy} + (1-x)J_{C,yy}, \\
&F_{xy} = J_A - J_B + xJ_{A,x} - xJ_{B,x} + yJ_{A,y} + (1-y)J_{B,y} - J_{C,y} + xyJ_{A,xy} \\
&+ x(1-y)J_{B,xy} + (1-x)J_{C,xy},
\end{aligned}$$

where $J_{i,x}$ and $J_{i,xx}$ are the first and second partial derivative of J_i with respect to x, respectively, and similarly for $J_{i,y}$, $J_{i,yy}$ and $J_{i,xy}$.

2. Using market-clearing condition for the endowment claim, plugging in the expression for agent C's optimal portfolio choice w_s^C from (24), and substituting for $(\mu - r)/\sigma^2$ from the expression

[59] For a general state space $x \in [x_L, x_H]$, we use a linear transformation $x_j = x_L + 0.5(x_H - x_L)(1 + \omega_j)$.

for $w_s^{A,*}$, we will get the first of the two equations that $w_s^{A,*}$ and w_s^B have to satisfy:

$$\begin{aligned}
1 &= xyw_s^{A,*} + x(1-y)w_s^B + (1-x)w_s^C \\
&= xyw_s^{A,*} + x(1-y)w_s^B \\
&+ (1-x)\frac{1}{\gamma_C}\left\{\frac{\mu - r}{\sigma^2} + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left[\frac{J_{C,x}}{J_C}x\left(yw_s^{A,*} + (1-y)w_s^B - 1\right) + \frac{J_{C,y}}{J_C}y(1-y)\left(w_s^{A,*} - w_s^B\right)\right]\right\} \\
&= xw_s^{A,*} + yw_s^B + (1-x)\frac{1}{\gamma_C}\left\{\gamma_A w_s^{A,*} - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left[\frac{J_{A,x}}{J_A}x\left(yw_s^{A,*} + (1-y)w_s^B - 1\right)\right.\right. \\
&\left.+\frac{J_{A,y}}{J_A}y(1-y)\left(w_s^{A,*} - w_s^B\right)\right] \\
&\left.+\left(\frac{1-\gamma_C}{1-\psi_C}\right)\left[\frac{J_{C,x}}{J_C}x\left(yw_s^{A,*} + (1-y)w_s^B - 1\right) + \frac{J_{C,y}}{J_C}y(1-y)\left(w_s^{A,*} - w_s^B\right)\right]\right\}
\end{aligned}$$

To get the second equation, I plug in the expression for $(\mu - r)/\sigma^2$ from A's optimal portfolio $w_s^{A,*}$ in the expression for w_s^B:

$$\begin{aligned}
w_s^B &= \frac{1}{\gamma_B}\left[\frac{\mu - r}{\sigma^2} + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}x\left(yw_s^A + (1-y)w_s^B - 1\right) + \frac{J_{B,y}}{J_B}y(1-y)(w_s^A - w_s^B)\right)\right] \\
&= \frac{1}{\gamma_B}\left[\gamma_A w_s^A - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}x\left(yw_s^A + (1-y)w_s^B - 1\right) + \frac{J_{A,y}}{J_A}y(1-y)(w_s^A - w_s^B)\right)\right. \\
&\left.+\left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}x\left(yw_s^A + (1-y)w_s^B - 1\right) + \frac{J_{B,y}}{J_B}y(1-y)(w_s^A - w_s^B)\right)\right]
\end{aligned}$$

We can rewrite the systems of equation as

$$\begin{aligned}
a_{11}w_s^{A,*} + a_{12}w_s^B &= b_1 \\
a_{21}w_s^{A,*} + a_{22}w_s^B &= b_2
\end{aligned}$$

where

$$\begin{aligned}
a_{11} &= xy + (1-x)\frac{1}{\gamma_C}\left[\gamma_A - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}xy + \frac{J_{A,y}}{J_A}y(1-y)\right)\right. \\
&\left.+\left(\frac{1-\gamma_C}{1-\psi_C}\right)\left(\frac{J_{C,x}}{J_C}xy + \frac{J_{C,y}}{J_C}y(1-y)\right)\right], \\
a_{12} &= x(1-y) + (1-x)\frac{1}{\gamma_C}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}x(1-y) - \frac{J_{A,y}}{J_A}y(1-y)\right)\right. \\
&\left.+\left(\frac{1-\gamma_C}{1-\psi_C}\right)\left(\frac{J_{C,x}}{J_x}x(1-y) - \frac{J_{C,y}}{J_C}y(1-y)\right)\right], \\
a_{21} &= \frac{1}{\gamma_B}\left[\gamma_A - \left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}xy + \frac{J_{A,y}}{J_A}y(1-y)\right) + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}xy + \frac{J_{B,y}}{J_B}y(1-y)\right)\right],
\end{aligned}$$

$$
\begin{aligned}
a_{22} &= -1 + \frac{1}{\gamma_B}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\left(\frac{J_{A,x}}{J_A}x(1-y) + \frac{J_{A,y}}{J_A}y(1-y)\right)\right. \\
&\quad \left. + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\left(\frac{J_{B,x}}{J_B}x(1-y) - \frac{J_{B,y}}{J_B}y(1-y)\right)\right], \\
b_1 &= 1 + (1-x)\frac{1}{\gamma_C}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\frac{J_{A,x}}{J_A}x + \left(\frac{1-\gamma_C}{1-\psi_C}\right)\frac{J_{C,x}}{J_C}x\right], \\
b_2 &= \frac{1}{\gamma_B}\left[-\left(\frac{1-\gamma_A}{1-\psi_A}\right)\frac{J_{A,x}}{J_A}x + \left(\frac{1-\gamma_B}{1-\psi_B}\right)\frac{J_{B,x}}{J_B}x\right].
\end{aligned}
$$

The system of equations above can be solved easily to get $w_s^{A,*}$ and w_s^B.

3. Since the return volatility can be written as

$$
\sigma = \frac{\sigma_D}{1 + \frac{F_x}{F}x\left[yw_s^A + (1-y)w_s^B - 1\right] + \frac{F_y}{F}y(1-y)\left(w_s^A - w_s^B\right)}, \tag{B.3}
$$

when the margin constrains for agent A bind, from equation (7) with $\nu = 1$, we must have

$$
w_s^{A,const} = \frac{1 - \frac{F_x}{F}x + \left(\frac{F_x}{F}x - \frac{F_y}{F}y\right)(1-y)w_s^B}{\alpha\sigma_D - \left[\frac{F_x}{F}x + \frac{F_y}{F}(1-y)\right]y} \tag{B.4}
$$

So, we have $w_s^A \leq w_s^{A,const}$. Then from (25) we can find A and B's portfolio weights in the risky asset

$$
w_s^A = \min\left(w_s^{A,*}, w_s^{A,const}\right),
$$

where $w_s^{A,const}$ is given in equation (B.4).

4. From stock market clearing, we can get C's optimal portfolio weight

$$
w_s^C = \frac{1 - xy\,w_s^A - x(1-y)\,w_s^B}{1-x}
$$

5. Using the expression for the return volatility in equation (19) and plugging in expressions for σ_x and σ_y from equations (16) and (17), the expression for return volatility is

$$
\sigma = \frac{\sigma_D}{1 + \frac{F_x}{F}x\left[yw_s^A + (1-y)w_s^B - 1\right] + \frac{F_y}{F}y(1-y)\left(w_s^A - w_s^B\right)}.
$$

6. Using the expression for σ above, state variable diffusions (σ_x and σ_y) can be found from equations (16) and (17):

$$
\sigma_x = \left[yw_s^A + (1-y)w_s^B - 1\right]\sigma, \quad \text{and} \quad \sigma_y = \left(w_s^A - w_s^B\right)\sigma.
$$

7. From the expression for w_s^C, σ, σ_x, and σ_y, the expected excess return (risk premium) on the

risky asset is

$$\mu - r = \gamma_C w_s^C \sigma^2 - \left(\frac{1-\gamma_C}{1-\psi_C}\right)\left(\frac{J_{C,x}}{J_C} x\sigma_x + \frac{J_{C,y}}{J_C} y(1-y)\sigma_y\right)\sigma$$

8. Using the optimal consumption-wealth ratios $c_i = J_i$, we can then compute drifts of the state variables μ_x and μ_y as

$$\begin{aligned}
\mu_x &= \left[y w_s^A + (1-y) w_s^B - 1\right](\mu - r - \sigma^2) - y J_A - (1-y) J_B + F \\
\mu_y &= \left(w_s^A - w_s^B\right)(\mu - r) - J_A + J_B - \left[y w_s^A + (1-y) w_s^B\right]\left(w_s^A - w_s^B\right)\sigma^2.
\end{aligned}$$

9. From equation (18) the expected return on the risky asset can be calculated

$$\begin{aligned}
\mu = {} & \mu_D + F - \frac{F_x}{F}\left[\kappa(\bar{x} - x) + x(\mu_x + \sigma_D \sigma_x)\right] - \frac{F_y}{F}\left[\kappa(\bar{y} - y) + y(1-y)(\mu_y + \sigma_D \sigma_y)\right] \\
& + \left[\left(\frac{F_x}{F}\right)^2 - \frac{1}{2}\frac{F_{xx}}{F}\right] x^2 \sigma_x^2 + \left[\left(\frac{F_y}{F}\right)^2 - \frac{1}{2}\frac{F_{yy}}{F}\right] y^2 (1-y)^2 \sigma_y^2 + \left[2\left(\frac{F_x}{F}\right)\left(\frac{F_y}{F}\right) - \frac{F_{xy}}{F}\right] xy(1-y)\sigma_x \sigma_y.
\end{aligned}$$

10. The real interest rate is

$$r = \mu - (\mu - r).$$

11. Plugging expressions above into agent i's HJB equations in (22), we get the residual functions for agent i:

$$\begin{aligned}
0 = {} & -(\rho + \kappa) + \frac{1}{\psi_i} J_i + \left(1 - \frac{1}{\psi_i}\right)\left[r + w_s^i (\mu - r) - \frac{\gamma_i}{2}\left(w_s^i\right)^2 \sigma^2\right] \\
& - \frac{1}{\psi_i}\left\{\frac{J_{i,x}}{J_i}\left[\kappa(\bar{x} - x) + x\mu_x\right] + \frac{J_{i,y}}{J_i}\left[\kappa(\bar{y} - y) + y(1-y)\mu_y\right] + (1-\gamma_i)\left(\frac{J_{i,x}}{J_i} x\sigma_x + \frac{J_{i,y}}{J_i} y(1-y)\sigma_y\right) w_s^i \sigma\right\} \\
& - \frac{1}{2\psi_i}\left[\left(\frac{\psi_i - \gamma_i}{1 - \psi_i}\right)\left(\frac{J_{i,x}}{J_i} x\sigma_x + \frac{J_{i,y}}{J_i} y(1-y)\mu_y\right)^2 + \frac{J_{i,xx}}{J_i} x^2 \sigma_x^2 + 2\frac{J_{i,xy}}{J_i} xy(1-y)\sigma_x \sigma_y + \frac{J_{i,yy}}{J_i} y^2 (1-y)^2 \sigma_y^2\right].
\end{aligned}$$

C Additional Model Results

C.1 Heterogeneous vs. Representative Intermediaries

I simulate representative- and heterogeneous-intermediary models for 3,000 quarters 20,000 times and examine the distribution of risk premium volatility. Figure G.1 in shows these distributions. As expected, the model with heterogeneous intermediaries exhibits more variation in risk premia than the one with a representative financial sector. Since the aggregate intermediary sectors in both models are (almost) identical, any excess variation in risk premia in the heterogeneous intermediary model has to be due to state variable y. In my calibration, approximately 20% of the variation in risk premia can be attributed to heterogeneity in the financial sector (state variable y). Therefore, failing to account for heterogeneity among intermediaries can lead to missing a substantial portion of the variation in risk premia.

C.2 Three-Dimensional Plots

Figure G.2 plots various objects the unconstrained equilibrium, where $\bar{\theta}_t = \bar{m}$. All variables are functions of the two state variables in the model: x (wealth share of agents A and B, i.e. the financial sector) and y (wealth share of A agents in the financial sector). These are the same objects plotted in solid blue line in Figures 2 and 3 but in three dimensions.D

Figure G.3 presents various variables in the equilibrium with time-varying margin constraint in the endowment model with $\bar{\theta}_t = \frac{1}{\alpha\sigma_t}$ as functions of state variables (x_t, y_t). These are the same equilibrium objects plotted in dashed red line in Figures 2 and 3 but in three dimensions.

D Internal Capital Market Regulation

Affiliate Transactions (Regulation W) Section 23A of the Federal Reserve Act (12 USC 371c) is the primary statute governing transactions between a bank and its affiliates. Section 23A (1) designates the types of companies that are affiliates of a bank; (2) specifies the types of transactions covered by the statute; (3) sets the quantitative limitations on a bank's covered transactions with any single affiliate, and with all affiliates combined; and (4) sets forth collateral requirements for certain bank transactions with affiliates.

Overview of Section 23A:

> *Section 23A prohibits a bank from initiating a "covered transaction" with an affiliate if, after the transaction, (i) the aggregate amount of the bank's covered transactions with that particular affiliate would exceed 10 percent of the bank's capital stock and surplus, or (ii) the aggregate amount of the bank's covered transactions with all affiliates would exceed 20 percent of the bank's capital stock and surplus.*[60]
>
> *Section 23A requires all covered transactions between a bank and its affiliate to be on terms and conditions consistent with safe and sound banking practices ("Safety and Soundness Requirement").*
>
> *Extensions of credit to an affiliate and guarantees, letters of credit, and acceptances issued on behalf of an affiliate ("credit transactions") must be secured by a statutorily defined amount of collateral, ranging from 100 to 130 percent of the covered transaction amount. Securities issued by an affiliate and low-quality assets are not acceptable collateral for any credit transaction with an affiliate. In addition, the attribution rule provides that any transaction by a bank with any person is deemed to be an affiliate transaction subject to section 23A to the extent that the proceeds of the transaction are used for the benefit of, or transferred to, an affiliate.*

Overview of Section 23B:

> *Section 23B requires that certain transactions, including all covered transactions, be on market terms and conditions ("Market Terms Requirement"). In addition to covered transactions, the Market Terms Requirement applies to: (i) any sale of assets by the bank to an affiliate; (ii) any payment of money or furnishing of services by the bank to an affiliate; (iii) any transaction in which an affiliate acts as agent or broker for the bank or any other person if the bank is a participant in the transaction; and (iv) any transaction by the bank with a third party if an affiliate has a financial interest in the third party or an affiliate is a participant in the transaction. In the absence of comparable transactions for identifying market terms, the bank must use terms, including credit standards that are at least as favorable to the bank as those that would be offered in good faith to nonaffiliated companies.*

[60] Covered transactions include loans and other extensions of credit to an affiliate, investments in the securities of an affiliate, purchases of assets from an affiliate, and certain other transactions that expose the bank to the risks of its affiliates.

Source: Federal Reserve Supervisory Policy and Guidance Topics, Affiliate Transactions (Regulation W).

E Data Sources

Broker-Dealer and Holding Company Data

Balance sheet data for broker-dealers and bank holding companies are from Tables L.130 and L.131 of Financial Accounts of the United States (Flow of Funds) from Federal Reserves, respectively. As noted in the description of Table L.130,

> *Security brokers and dealers are firms that buy and sell securities for a fee, hold an inventory of securities for resale, or do both. The firms that make up this sector are those that submit information to the Securities and Exchange Commission on one of two reporting forms, either the Financial and Operational Combined Uniform Single Report of Brokers and Dealers (FOCUS) or the Report on Finances and Operations of Government Securities Brokers and Dealers (FOGS). The major assets of the sector are collateral repayable from funding corporations in connection with securities borrowing (included in miscellaneous assets), debt securities and equities held for redistribution, customers' margin accounts, and security repurchase agreements (reverse repos). Sector operations are financed largely by net transactions with parent companies, customers' cash accounts, loans for purchasing and carrying securities from depository institutions, and security repurchase agreements.*

Also from Table L.131's description for holding companies,

> *. . . the holding companies sector consists of all top-tiered bank holding companies, savings and loan holding companies, U.S. Intermediate Holding Companies (IHCs), and securities holding companies (collectively "holding companies") that file the Federal Reserve's Form FR Y-9LP, Parent Company Only Financial Statements for Large Holding Companies, FR Y-9SP, Parent Company Only Financial Statements for Small Holding Companies, or FR 2320, Quarterly Savings and Loan Holding Company Report. Holding companies required to file FR Y-9LP include those with total consolidated assets of $1 billion or more or meet other criteria, such as having a material amount of debt or equity securities outstanding that are registered with the Securities and Exchange Commission, being engaged in significant nonbanking activity, or conducting off-balance-sheet activities either directly or through a nonbank subsidiary. Those holding companies required to file FR Y-9SP have total consolidated assets less than $1 billion. Form FR 2320 must be filed by top-tier savings and loan holding companies exempt from initially filing the Y-9LP or Y-9SP, because even though they own a savings and loan institution, that is not their primary line of business. Mutual stock companies that file the FR 2320 are excluded because they do not hold any assets or liabilities at the holding company level. The major assets of holding companies, other than small amounts of loans and securities, are net transactions with their subsidiaries; this includes equity investments in subsidiaries and associated banks and net balances due from subsidiaries and related depository institutions. The main source of funding for the sector is the issuance of corporate bonds and commercial paper.*[61]

[61] The holding companies sector has a large increase in the level of assets and liabilities in the 2009:Q1 because a number of large financial institutions became bank holding companies. These companies (including Goldman Sachs, Morgan Stanley, American Express, CIT Group, GMAC, Discover Financial Services, and IB Finance) had not previously been included in the financial accounts.

Test Assets

Test assets for time-series and cross-sectional asset pricing tests are from two sources: (i) equity portfolios (25 portfolios formed on size and book-to-market and 10 momentum portfolios) are from Ken French's Data Library, and (ii) non-equity assets are from HKM obtained from Asaf Manela's website and include 10 maturity-sorted US government and 10 corporate bond portfolios sorted on yield spreads, 6 sovereign bond portfolios based on a two-way sort on a bond's covariance with US equity market and bond's S&P rating, 54 portfolios of S&P 500 index options sorted on moneyness and maturity split by contract type (27 calls and 27 puts), 20 CDS portfolios sorted by spreads using single-name 5-year contracts, 23 commodity portfolios with at least 25 years of return data, and 12 foreign exchange currency portfolios, six sorted on interest rate differentials and six sorted on momentum. Except for Treasury bond portfolios which are from CRSP, non-equity test assets in HKM are from previous studies.

Intermediary Asset Pricing Factors

AEM and HKM factors are from Tyler Muir's and Asaf Manela's websites, respectively. AEM leverage factor is defined as the seasonally adjusted growth rate in broker-dealer book leverage from Table L.130 of the Flow of Funds, where leverage is defined as total financial assets divided by total financial assets minus total volatility. The intermediary capital ratio in HKM is the ratio of total market equity to total market assets (book debt plus market equity) of primary dealer holding companies of the New York Fed. Shocks to capital ratio (HKM capital factor) are defined as AR(1) innovations in the capital ratio, scaled by the lagged capital ratio. Data for publicly-traded holding companies of primary dealers are from CRSP/Compustat and Datastream. Primary dealers are large and sophisticated institutions and serve as trading counterparties of the NY Fed in its implementation of monetary policy. For the current and historical list of primary dealers see this link.

F Robustness Checks for Empirical Results

F.1 Predictive Regressions

Exclude the Great Recession form the Sample

As a robustness check, I remove the Great Recession (years 2007 to 2009) form the sample and rerun the predictive regressions in equation (31) with the market excess return as the dependent variable. Table G.1 presents the results. Consistent with the first two columns of Table 4 with the full sample, we also see negative and significant coefficient γ_y with additional predictive power over control variables in the sample excluding the 2008 financial crisis. Importantly, the predictive regression results are robust to excluding the Great Recession from the sample: It is not just the financial crisis that drives my predictability results.

Using the Cyclical Component of y as Predictor

One might be concerned about time-series trends in state variable y impacting the predictive regression results. Figure G.4 presents the trend and cyclical components of y^{data}, defined in equation (30). Hodrick-Prescott filter with smoothing parameter of 1600 is used to separate the time series into trend and cyclical components. We observe that the cyclical component shows procyclical behavior particularly in the late part of the sample.

Table G.2 presents the results. Consistent with the first two columns of Table 4, we also see negative and significant coefficient γ_y for the lagged cyclical component of state variable y (y^{cyc}) with additional predictive power over control variables. I also considered other ways to detrend the time series. The predictive regression results are robust to excluding the time trends in state variable y and using only its cyclical components as the main forecasting variable. In unreported regression, the forecasting regression is also robust to using 1-year, and 5-year growth rates, as well as the AR(1) residual of y.

Include Factors from Representative Intermediary-Based Models

As robustness, in Table G.3, I further examine the predictive power of the composition of financial intermediaries (captured by state variable y) in the presence of factors from *representative* intermediary asset pricing models studied in AEM and HKM. That is, I include AEM and/or HKM factors in predictive regressions in equation (31):

$$R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y \, y_t + \gamma^i_{\text{Rep}} \, \text{Rep}_t + \gamma^i_{\text{Ctrl}} \, \text{Ctrl}_t + \varepsilon_{t+1\to t+4},$$

where Rep represents the vectors of representative intermediary factors: broker-dealer leverage from AEM and BHC capital ratio from HKM. Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: wealth share of the aggregate financial sector (x from the model defined in 29), fluctuations in the aggregate consumption-wealth ratio (*cay* variable) defined in Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev et al. (2009). For reference, the first column repeats the regression in Column (1) of Table 4. In Column (2), HKM's intermediary capital ratio (CapRatio) is added as an additional predictor. We observe that the coefficient on CapRatio is not statistically significant and the R^2 is only slightly increased (from 0.43 to 0.45). Removing y in Column (3) substantially reduces R^2 by 13%, emphasizing the predictive power of my measure of intermediary heterogeneity beyond CapRatio. In Column (4), AEM's borker-dealer leverage (BDLev) is added as an additional predictor. Similarly, the coefficient on BDLev is not statistically significant and the R^2 is only slightly increased (from 0.43 to 0.47). Removing y in Column (5) however, does not substantially reduce R^2 (only by 2%). Finally, in the last column, all three predictors are included simultaneously: The coefficient on y remains negative and highly significant with a large R^2 of 0.47. The coefficient is also economically significant: a 1% decrease in wealth share of dealers in the financial sector predicts a 1.2% (quarterly, 4.8% annualized) increase in market risk premium over the next four quarters. As before, I include several control variables that are known in the literature to forecast returns: Wealth share of the aggregate financial sector (x from the model defined in 29), fluctuations in the aggregate consumption-wealth ratio (*cay* variable) defined in Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev et al. (2009).

F.2 Cross-Sectional Asset Pricing Tests

HIFac's Pricing Performance

Figure G.5 visually shows the HIFac's pricing performance: The top panel plots the annualized realized against the predicted excess returns for the 55 equity and bond portfolios when HIFac is the only pricing factor (Column (1) in Table 5). Most of the portfolios line up closely to the 45-degree line. The bottom panel is similar to the top panel when HIFac and AEM are used as pricing factors, corresponding to Column (4) in Table 5. The model slightly outperforms the one in panel (a) as shown in above.

One-Way Sorted CRSP Portfolios

To empirically verify the positive price of risk for innovations in the wealth share of dealers in the financial sector, I sort stocks based on their exposures to these shocks and form portfolios by quintiles on a 10-year trailing window. I consider all common stocks (share codes 10 and 11) in the CRSP universe from Amex, NASDAQ, and NYSE (exchange codes 1,2, and 3). For every stock i at quarter t, I regress its quarterly excess return on constant and innovations in the heterogeneous intermediary factor (HIFac), defined in equation (32):

$$R^e_{i,t} = \alpha_i + \beta_{i,\text{HIFac}}\, \text{HIFac}_t + \xi_{i,t}$$

The coefficient $\beta_{i,\text{HIFac}}$ measures the exposure of firm i's stock to the factor's innovations. I then sort stocks into quintiles every quarter according to their $\beta_{i,\text{HIFac}}$.

The average returns of the beta-sorted portfolios are reported in Table G.4, along with return volatilities, average book-to-market ratio, average market cap, and alphas from CAPM and Fama-French three-factor model. Consistent with model's implications, when sorted on β_{HIFac}, average risk premia are increasing from the portfolio of low-beta stocks to the high-beta quintile. Excess returns are monotonically increasing from quintile one to five and the top portfolio earns an approximately 5% premium over the lowest quintile.

Two-Way Sorted CRSP Portfolios

In this section, I verify the results above are robust to double-sorting with asset pricing factors from recent models with representative intermediaries. In this exercise, I independently double-sort CRSP stocks into three-by-three portfolios on their exposures to the heterogeneous intermediary factor (HIFac) and either AEM or HKM representative intermediary asset pricing factors. Table G.5 reports returns for double-sorted portfolios on exposures to HIFac and AEM and HKM betas. The return spread on HIFac-beta-sorted portfolios is 4.34% and 3.14% per year among stocks with low exposures to the AEM leverage and HKM capital factors, respectively.

This exercise demonstrates that the heterogeneity in the financial sector is an important risk factor and has pricing information above and beyond representative intermediary asset pricing factors in AEM and HKM: even within portfolios sorted based on AEM or HKM factor betas, I see a monotonic progression in returns from low- to high-HIFac beta portfolios.

The Heterogeneous Intermediary Factor-Mimicking Portfolio

As emphasized above, the main argument of the paper is that the heterogeneity in the intermediary sector has important implication for asset prices. To conduct additional robustness tests, in this section, I project the heterogeneous intermediary factor (HIFac) onto the space of traded returns to form a factor-mimicking portfolio that mimics the HIFac. To further verify that this heterogeneity an important source of risk, I evaluate the heterogeneous intermediary factor-mimicking portfolio (HIMP) relative to the mimicking portfolios for representative intermediary factors in AEM and HKM. I show that the mimicking portfolios for these representative intermediary factors cannot fully span the HIMP and there is more to be captured by the heterogeneity within the financial sector.

This approach also allows me to run tests using higher frequency data and longer time series. Moreover, since the mimicking portfolio is a traded excess return, I can evaluate the model by testing alphas in the time-series regression without the need to estimate the cross-section risk prices.

Construction of HIMP To construct mimicking portfolio of the heterogeneous intermediary factor (HIFac), I follow AEM and project this factor, onto the space of excess returns by running the following regression:

$$\mathrm{HIFac}_t = a_{\mathrm{HI}} + b'_{\mathrm{HI}} \left[BL, BM, BH, SL, SM, SH, Mom, Bond\right]_t + \varrho_t, \tag{F.1}$$

where HIFac is the heterogeneous intermediary factor defined in equation (32), and BL, BM, BH, SL, SM, SH are, respectively, the excess returns of the six Fama-French portfolios on size (Small and Big) and book-to-market (Low, Medium, and High), and Mom is the momentum factor, obtained from Ken French's data library. $Bond$ is the first principal component (PC) of excess returns on six Treasury bond portfolios sorted by maturity from CRSP. The heterogeneous intermediary mimicking portfolio (HIMP) is then given by

$$\mathrm{HIMP}_t = \widetilde{b}'_{\mathrm{HI}} \left[BL, BM, BH, SL, SM, SH, Mom, Bond\right]_t, \tag{F.2}$$

where $\widetilde{b}_{\mathrm{HI}} = \frac{b'_{\mathrm{HI}}}{\sum b_{\mathrm{HI}}} = [-0.34, 0.20, -1.04, -0.09, 0.41, 1.64, 1.04, -0.83]$ positively loading on the momentum factor.

HIMP vs. Mimicking Portfolios for AEM and HKM Factors To further verify that my heterogeneous intermediary factor captures sources of risk beyond the factors from representative intermediary asset pricing models, in this section I evaluate the performance of HIMP with mimicking portfolios for AEM and HKM factors. I similarly construct mimicking portfolios for AEM's broker-dealer leverage and HKM's holding company capital factors using quarterly data for the two factors from Tyler Muir's and Asaf Manela's websites, respectively.[62] The mimicking portfolio for the heterogeneous intermediary factor has Sharpe ratio of 0.45 over the sample period (1970Q1 to 2017Q3), much higher than Sharpe ratios for AEM and HKM factor-mimicking portfolios (0.21 and 0.27, respectively).

[62] The loadings for AEM and HKM factor-mimicking portfolios are $\widetilde{b}_{\mathrm{AEM}} = [-0.98, 0.50, -0.03, -0.26, 0.96, 0.05, 0.16, 0.59]$ and $\widetilde{b}_{\mathrm{HKM}} = [0.30, 0.03, 0.58, -0.06, -0.16, 0.25, 0.09, -0.03]$.

To evaluate the importance of heterogeneity in the financial sector above an beyond representative intermediary factors, I regress HIFac on mimicking portfolios for AEM and HKM factors in the following regression:

$$\mathrm{HIMP}_t = \alpha_{\mathrm{MP}} + \beta_{\mathrm{FMP}}' \, \mathrm{FMP}_t + \epsilon_t, \tag{F.3}$$

where FMP is either the mimicking portfolio for broker-dealer leverage factor from AEM (AEM_MP), or the mimicking portfolio for capital factor for primary dealers' holding companies from HKM (HKM_MP), or both AEM_MP and HKM_MP. Notice the mimicking portfolios are traded excess returns, thus I can evaluate the model by testing alphas in the time-series regression without the need to estimate the cross-section risk prices. If HIMP is fully "explained" by AEM_MP, HKM_MP, or both, I expect to see small and insignificant α_{MP} in the regression above. I find the opposite to be true, however.

Table G.6 presents the results. In Columns 1 and 2, I run univariate regression where the dependent variables are AEM_MP and HKM_MP, respectively. In both cases the intercept, α_{MP} is statistically significant at 1% level and the R^2 of the regressions are relatively low at 0.14 and 0.32, respectively. In Column 3, I added value-weighted return from CRSP (MktRF) to HKM_MP as independent variables which leads to very similar results to Column 3. In Column 4, I add both AEM and HKM factor-mimicking portfolios as right-hand-side variables in equation (F.3). We observe a large and significant α_{MP} and relatively small R^2. Adding MktRF in Column 5 to the regression in Column 4, further strengthen the results.

As a robustness check, I build factor-mimicking portfolios by projecting them instead onto the Fama-French three factors, the momentum factor, and the first PC of bond portfolios, and repeat the regressions in Table G.6. I arrive at very similar results: time-series alphas are large and significantly different from zero with low R^2 in all regressions. See Table G.7 in Appendix F.

This exercise confirms my earlier results: the heterogeneity in the financial sector is an important risk factor and has pricing information above and beyond representative intermediary asset pricing factors in AEM and HKM.

Alternative Projections for Factor-Mimicking Portfolios (FMPs) In this section, I repeat the exercise in Section F.2 with an alternative set of returns. I construct a mimicking portfolio for the heterogeneous intermediary factor (HIFac) by projecting it onto the space of excess returns running the following regression:

$$\mathrm{HIFac}_t = a_{\mathrm{HI}} + b_{\mathrm{HI}}' \left[\mathrm{MktRF}, SMB, HML, Mom, Bond\right]_t + \varrho_t, \tag{F.4}$$

where HIFac is the heterogeneous intermediary factor defined in equation (32), MktRF, SMB, and HML are the Fama-French three factors, Mom is the momentum factor, and $Bond$ is the first principal component (PC) of excess returns on six Treasury bond portfolios sorted by maturity from CRSP. The heterogeneous intermediary mimicking portfolio (HIMP) is then given by

$$\mathrm{HIMP}_t = \widetilde{b}_{\mathrm{HI}}' \left[\mathrm{MktRF}, SMB, HML, Mom, Bond\right]_t, \tag{F.5}$$

where $\widetilde{b}_{\mathrm{HI}} = \frac{b_{\mathrm{HI}}'}{\sum b_{\mathrm{HI}}} = [0.19, 0.51, 0.22, 0.29, -0.22]$ positively loading on the momentum factor.

I similarly construct mimicking portfolios for AEM's broker-dealer leverage and HKM's holding company capital factors using quarterly data for the two factors from Tyler Muir's and Asaf Manela's websites, respectively. The mimicking portfolio for the heterogeneous intermediary factor has Sharpe ratio of 0.43 over the sample period (1970Q1 to 2017Q3), much higher than Sharpe ratios for AEM and HKM factor-mimicking portfolios (0.24 and 0.28, respectively).

To evaluate the importance of heterogeneity in the financial sector above an beyond representative intermediary factors, I regress HIFac on mimicking portfolios for AEM and HKM factors in the following regression:

$$\mathrm{HIMP}_t = \alpha_{\mathrm{MP}} + \beta'_{\mathrm{FMP}}\,\mathrm{FMP}_t + \epsilon_t, \tag{F.6}$$

where FMP is either the mimicking portfolio for broker-dealer leverage factor from AEM (AEM_MP), or the mimicking portfolio for capital factor for primary dealers' holding companies from HKM (HKM_MP), or both AEM_MP and HKM_MP. If HIMP is fully "explained" by AEM_MP, HKM_MP, or both, I expect to see small and insignificant α_{MP} in the regression above. I find the opposite to be hold in the data, however. Table G.7 presents the results. Time-series alphas are positive and significant at 1% level in all columns similar to the regressions in Table G.6 from the main text.

G Appendix Figures and Tables

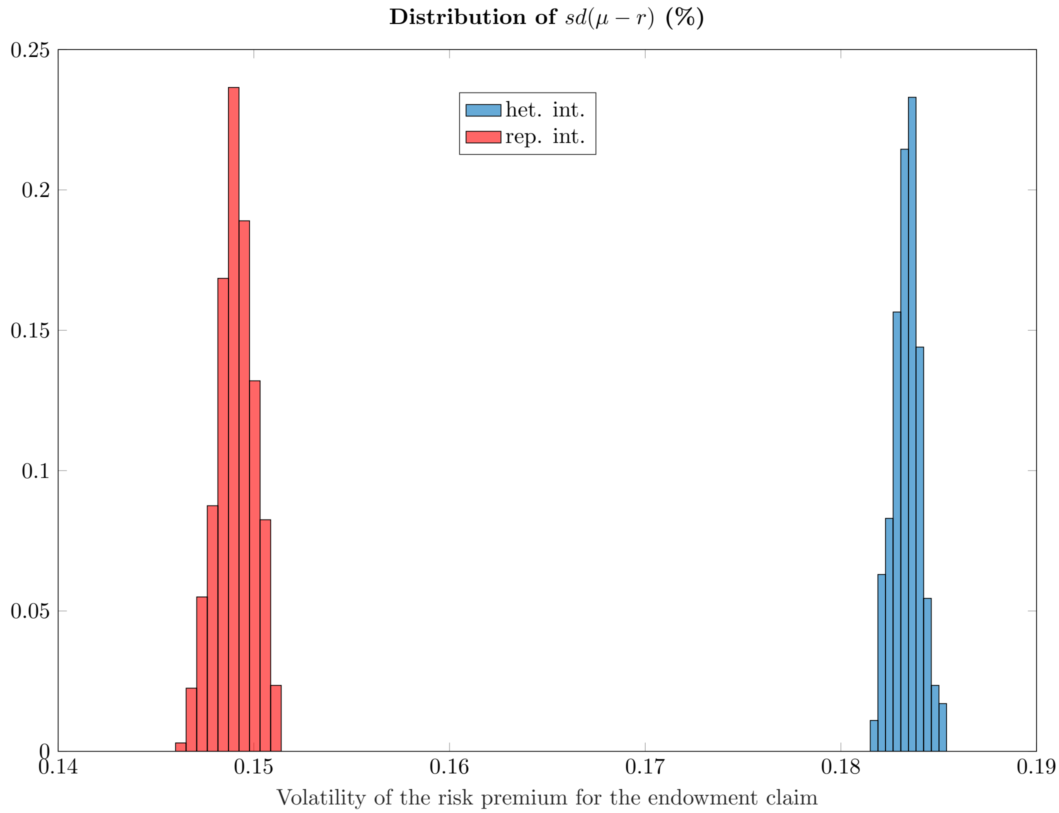


Figure G.1. Heterogeneous and representative intermediaries. This figure presents distribution of risk premia volatility in models with representative (red) and heterogeneous (blue) intermediaries. I simulate each model 20,000 times for 3,000 quarters. Notice that horizontal axis is the volatility of the risk premium for the endowment claim, which has relatively low volatility ($\sigma_D = 3.5\%$) relative to the market (approximately 16%). Therefore, equity premium volatility implied by the model is about five to six times larger than that of the endowment claim (approximately 1% for the heterogeneous intermediary model, for example).

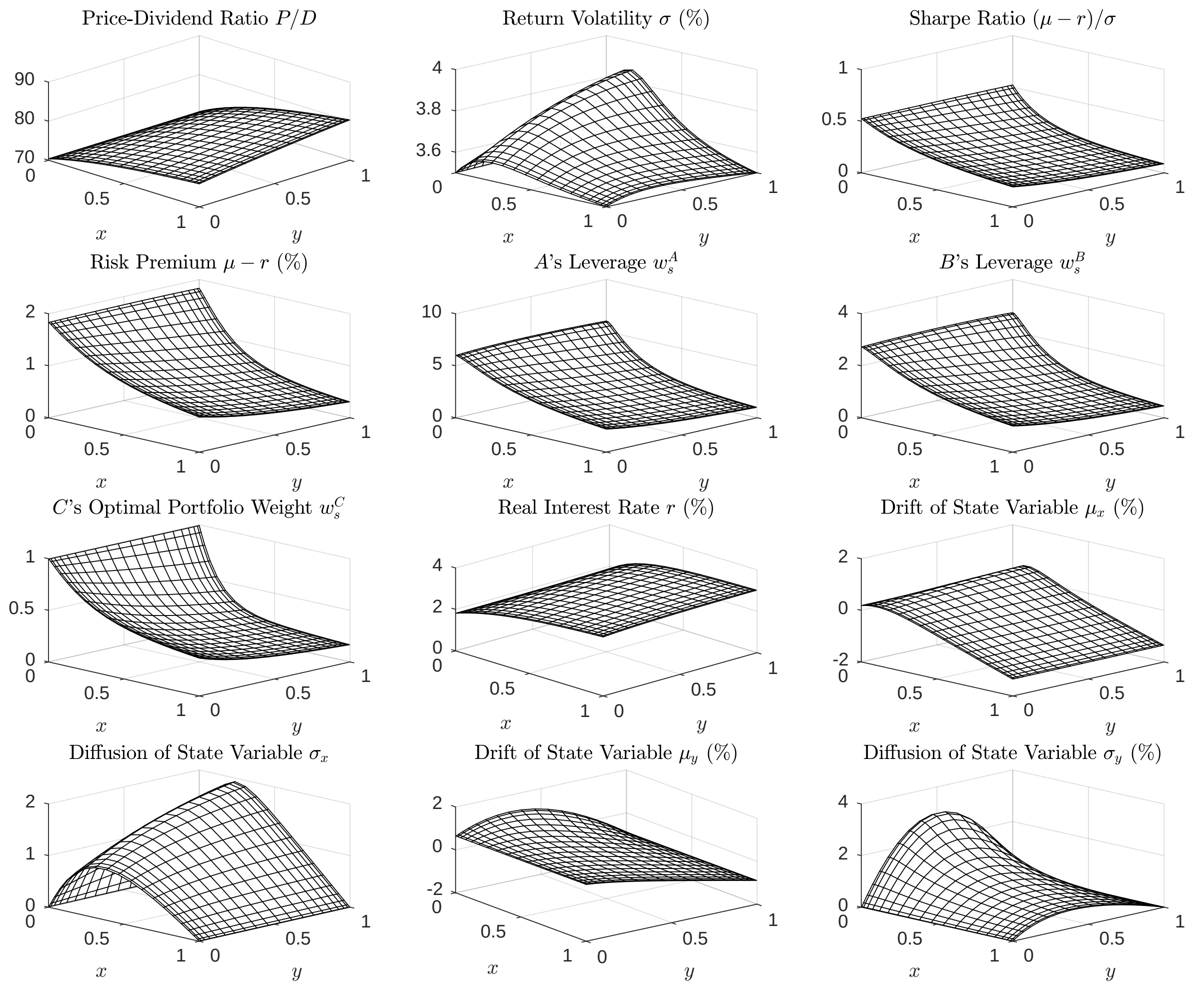


Figure G.2. Equilibrium in the unconstrained economy. This figure presents price-dividend ratio $1/F$, return volatility σ, Sharpe ratio and risk premium on the endowment claim, optimal portfolio weights of each type of agent (w_s^A, w_s^B, and w_s^C) as well as the real interest rate r_t and the drift and diffusion of state variables x and y (μ_x, σ_x, μ_y, and σ_y, respectively) in the *frictionless* economy as functions of state variable x (wealth share of the financial sector i.e. type A and B agents) and y_t (wealth share of type A agents in the financial sector) under the benchmark parameters in Table 1.

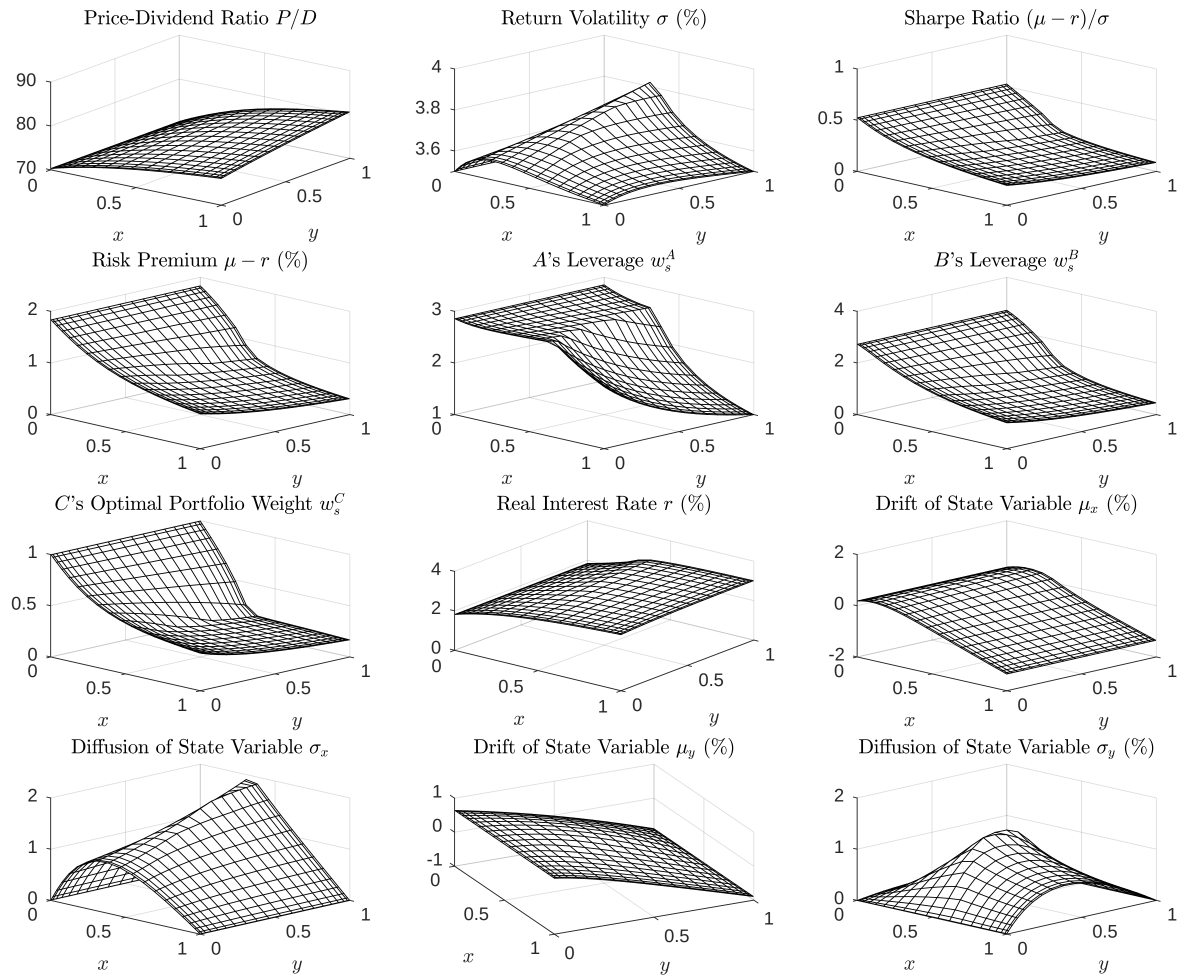
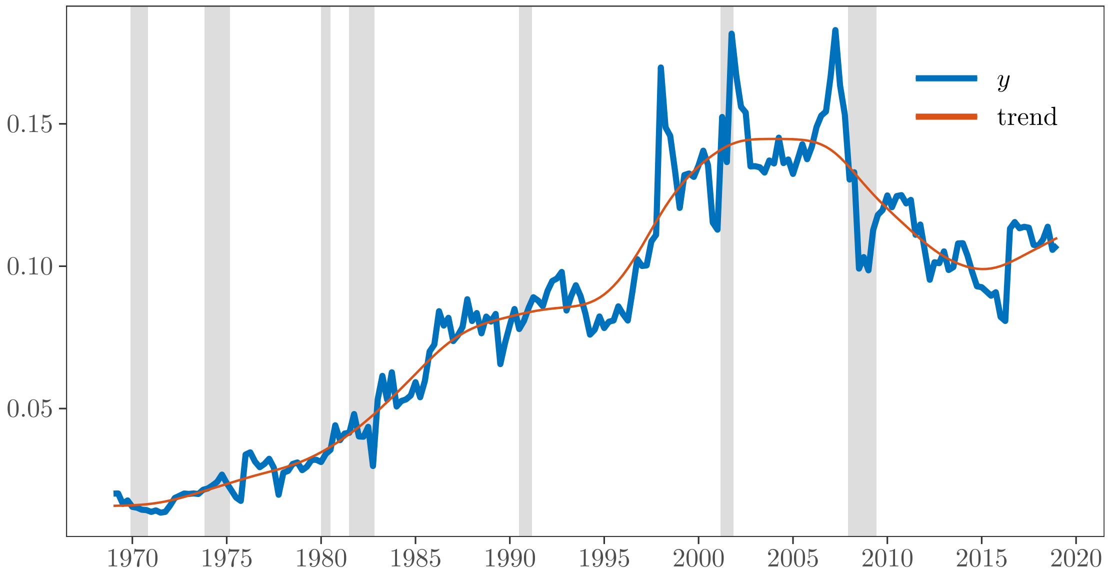


Figure G.3. Equilibrium in the economy with time-varying margin constraints. This figure presents price-dividend ratio $1/F$, return volatility σ, Sharpe ratio and risk premium on the endowment claim, optimal portfolio weights of each type of agent (w_s^A, w_s^B, and w_s^C) as well as the real interest rate r_t and the drift and diffusion of state variables x and y (μ_x, σ_x, μ_y, and σ_y, respectively) in the economy with *time-varying margin constraints* as functions of state variable x (wealth share of the financial sector i.e. type A and B agents) and y_t (wealth share of type A agents in the financial sector) under the benchmark parameters in Table 1.



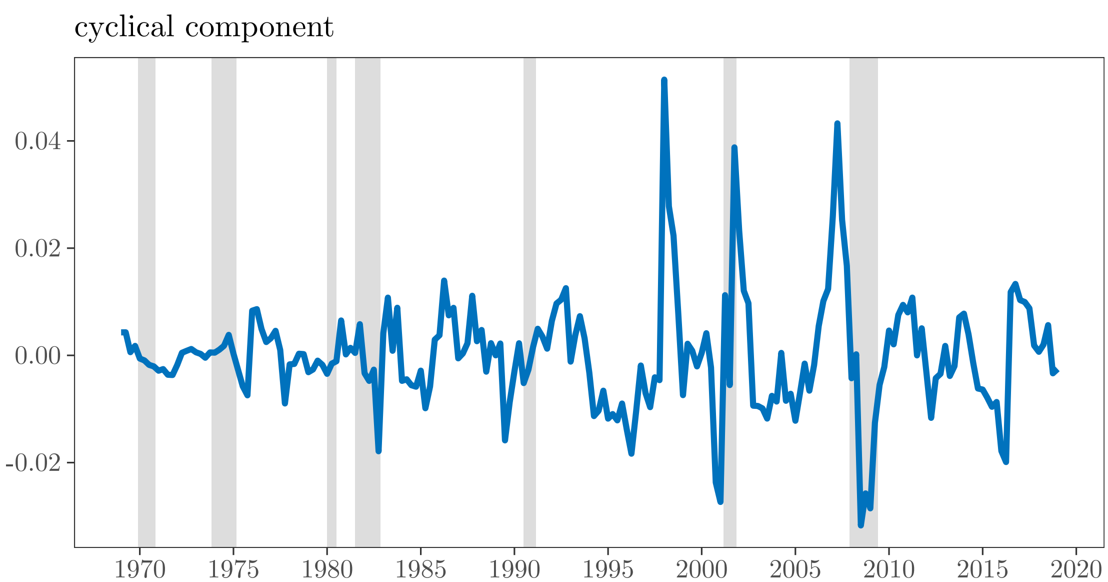
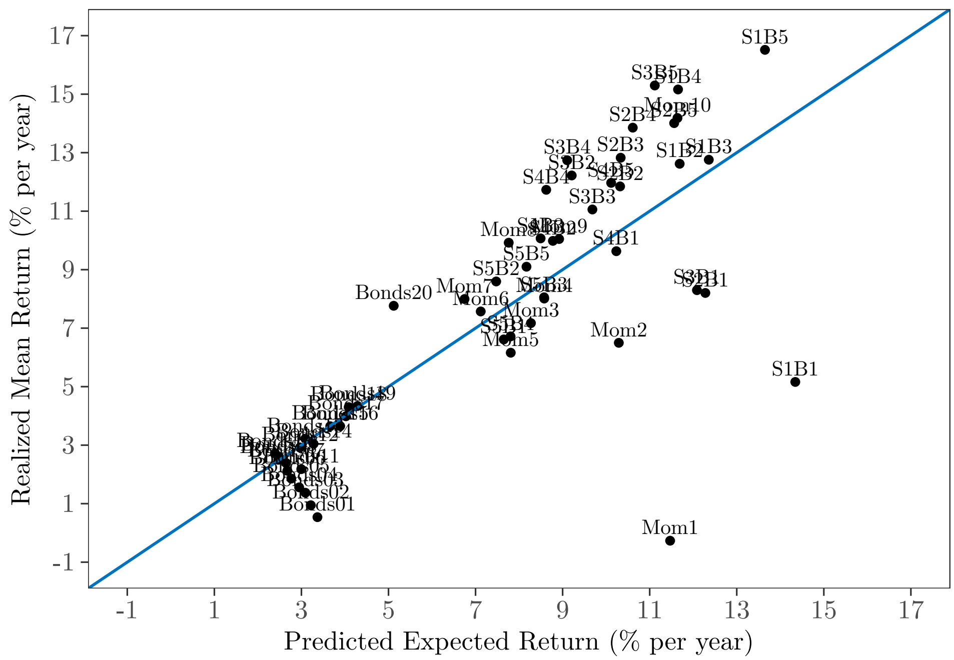
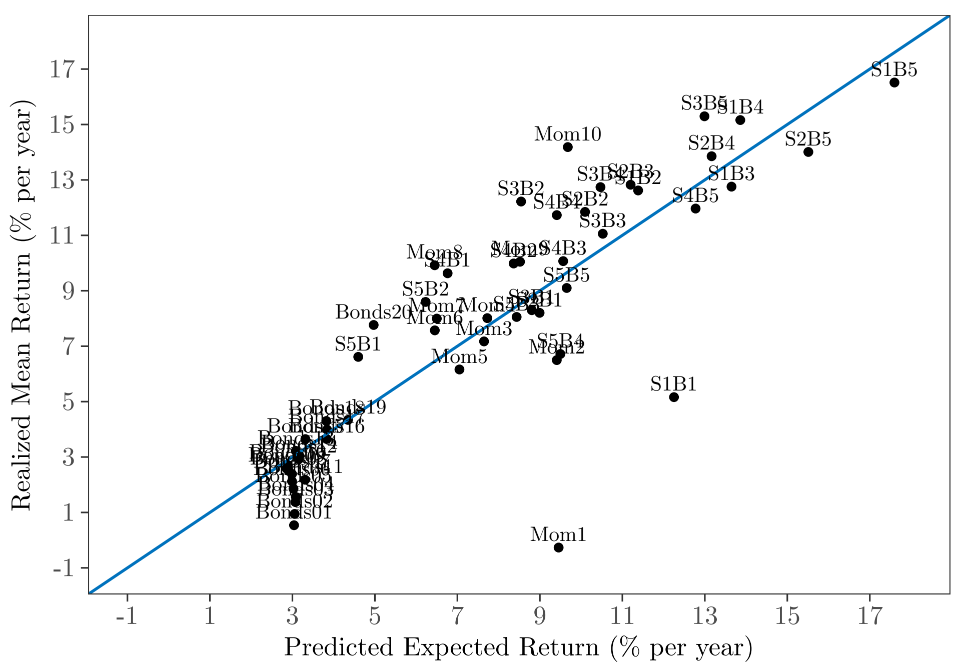


Figure G.4. Trend and cyclical components of state variable y in the data. This figure presents the trend and cyclical components of quarterly book equity share of the broker-dealers in the financial sector, y^{data}, defined in equation (30). Hodrick-Prescott filter with smoothing parameter of 1600 is used to separate the time series into trend and cyclical components. Book equity for BDs and commercial banks are computed from the Flow of Funds Tables L.130 and L.110, respectively. Sample period is from 1970Q1 to 2018Q4. The vertical shaded bars indicate NBER recessions.



(a) HIFac only.



(b) HIFac + AEM.

Figure G.5. Realized versus predicted mean returns: intermediary heterogeneity factor. This figure presents the realized mean excess returns of 35 equity portfolios (25 size and book-to-market-sorted portfolios and 10 momentum-sorted portfolios) and 10 Treasury bond portfolios (sorted by maturity), and 10 US corporate bond portfolios (sorted by yield spread) against the mean excess returns predicted by the single heterogeneous intermediary risk factor when only the heterogeneous intermediary factor (HIFac) (panel a) and HiFac and AEM factors (panel b) are used as pricing factor, respectively. The sample is quarterly from 1970Q1 to 2017Q4. Returns are reported in percent per year (quarterly percentages multiplied by four).

Table G.1. Predictive regressions: Excluding the Great Recession.
This table provides results for one-year ahead predictive regressions according to $R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y y_t + \gamma^i_{\text{Ctrl}} \text{Ctrl}_t + \varepsilon^i_{t+1\to t+4}$, using lagged equity share of broker-dealers in the financial sector, y^{data}_t defined in (30). Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: wealth share of the aggregate financial sector (x from the model defined in 29), *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings ratios (CAPE) from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variables are excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}). Broker-Dealer leverage is calculated using data from Table L.130 of Flow of Funds and is defined: Total Financial Assets/(Total Financial Assets − Total Liabilities). The sample is quarterly from 1974Q1 to 2018Q4. The Great Recession (years 2007–2009) is excluded form the sample. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable:	
	Mkt	
	(1)	(2)
y_t		-1.06^{***}
		(0.34)
Constant	-0.12	-0.13
	(0.14)	(0.10)
Controls	Yes	Yes
Observations	97	97
R^2	0.48	0.54
Adjusted R^2	0.45	0.51
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table G.2. Predictive regressions: Using cyclical component of y as predictor.
This table provides results for one-year ahead predictive regressions according to $R^i_{t+1\to t+4} - r^f_t = \gamma^i_0 + \gamma^i_y\, y^{cyc}_t + \gamma^i_{\mathrm{Ctrl}}\,\mathrm{Ctrl}_t + \varepsilon^i_{t+1\to t+4}$, using lagged *cyclical component* of equity share of broker-dealers in the financial sector, y^{cyc}_t as the main predictor. Hodrick-Prescott filter with smoothing parameter of 1600 is used to separate the time series of y into trend and cyclical components. Ctrl represents the vector of control variables that are known in the literature to forecast returns. I use the following control variables: wealth share of the aggregate financial sector (x from the model defined in 29), *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings ratios (CAPE) from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variables are excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}). Broker-Dealer leverage is calculated using data from Table L.130 of Flow of Funds and is defined: Total Financial Assets/(Total Financial Assets − Total Liabilities). The sample is quarterly from 1970Q1 to 2018Q4. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable:	
	Mkt	
	(1)	(2)
y^{cyc}_t		-4.47^{***}
		(1.06)
Constant	0.04	-0.06
	(0.19)	(0.16)
Controls	Yes	Yes
Observations	103	103
R^2	0.33	0.42
Adjusted R^2	0.29	0.39
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table G.3. Predictive regressions for the MKT excess return: Robustness
This table provides results for one-year ahead predictive regressions using lagged equity share of broker-dealers in the financial sector, the empirical proxy for state variable y defined in equation (30) which captures the composition of the financial sector, as well as intermediary equity capital ratio (from HKM) and leverage of broker-dealers (from AEM) as predictors of interest. I use the same control variables as in Table 4: wealth share of the aggregate financial sector (x from the model defined in 29), *cay* variable from Lettau and Ludvigson (2001), real price-dividend (PD) and cyclically adjusted price-earnings (CAPE) ratios from Robert Shiller's website, and variance risk premium (VRP) from Bollerslev et al. (2009). The dependent variable is excess holding period returns from quarter $t+1$ to quarter $t+4$ on the CRSP value-weighted portfolio (Mkt_{t+1}). The sample quarterly from 1974Q1 to 2017Q3. Broker-Dealer leverage is calculated using data from Table L.130 of Flow of Funds and is defined: Total Financial Assets/(Total Financial Assets − Total Liabilities). The capital ratio for New York Fed's primary dealer holding companies are downloaded from Asaf Manela's website. Hodrick (1992) standard errors are reported in parentheses to adjust for the fact that overlapping quarterly observations are used to forecast annual returns.

	Dependent variable: Mkt_{t+1}					
	(1)	(2)	(3)	(4)	(5)	(6)
y_t	-1.80^{***}	-1.98^{***}		-0.97^{*}		-1.19^{***}
	(0.43)	(0.48)		(0.51)		(0.43)
CapRatio_t		3.31	1.22			1.68
		(3.58)	(4.58)			(2.25)
BDLev_t				-1.35	-1.98^{***}	-1.13^{*}
				(0.85)	(0.56)	(0.59)
Constant	0.01	0.18	0.10	0.05	0.09	0.13
	(0.13)	(0.26)	(0.38)	(0.15)	(0.17)	(0.22)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	109	109	109	109	109	109
R^2	0.46	0.49	0.32	0.51	0.48	0.51
Adjusted R^2	0.43	0.45	0.28	0.47	0.45	0.47
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table G.4. One-way sorted CRSP portfolios on exposures to the heterogeneous intermediary factor. This table reports average excess returns, alphas, volatility, average book-to-market ratio, and average market capitalization for portfolios formed on their exposure to shocks to dealer wealth share in the financial sector. Shocks to dealer wealth share (HIFac) are defined as AR(1) innovations in the wealth share, scaled by the lagged wealth share as shown in equation (32). Data is quarterly from 1970Q1 to 2017Q3. Returns, volatilities, and alphas are annualized.

	L (1)	(2)	(3)	(4)	H (5)	HML (6)
Average Excess Return (%)	11.66	11.53	12.81	14.34	16.65	4.98
Volatility (%)	19.69	19.08	21.76	26.41	35.53	26.72
β_{HIFac}	−0.20	0.33	0.69	1.19	2.21	2.41
t-stat	−0.89	1.50	2.77	4.07	5.89	10.67
α_{CAPM}	4.77	4.33	4.56	4.75	4.44	−0.32
t-stat	3.40	3.90	3.63	2.84	1.57	−0.10
α_{FF3}	3.89	2.88	3.36	3.98	4.02	0.14
t-stat	3.14	4.21	5.45	5.04	2.24	0.05
Average Market Cap ($bn)	5.28	3.66	2.40	1.97	0.89	–

Table G.5. Two-way sorted CRSP portfolios.
This table reports average excess returns for portfolios independently double-sorted on their exposure to shocks to dealer wealth share in the financial sector (HIFac) and beta to the AEM leverage factor (AEM LevFac), as well as, double-sorted portfolios on HIFac beta and HKM capital ratio factor (HKM CapFac) beta. Shocks to dealer wealth share (HIFac) are defined as AR(1) innovations in the wealth share, scaled by the lagged wealth share as shown in equation (32). AEM leverage and HKM capital factors are from Tyler Muir's and Asaf Manela's websites, respectively. Returns are annualized in percentage points. Data is quarterly from 1970Q1 to 2017Q3.

			HIFac		
AEM LevFac	(1)	(2)	(3)	(3)−(1)	t-stat
(1)	12.85	14.05	17.19	4.34	1.49
(2)	11.47	12.75	14.63	3.16	1.25
(3)	11.42	12.19	14.67	3.24	1.25
(3)−(1)	−1.43	−1.86	−2.54	–	–
t-stat	−0.65	−0.85	−0.91	–	–
			HIFac		
HKM CapFac	(1)	(2)	(3)	(3)−(1)	t-stat
(1)	10.05	11.39	13.19	3.14	1.20
(2)	11.91	12.48	14.82	2.91	1.28
(3)	14.75	15.17	17.74	2.99	1.13
(3)−(1)	4.70	3.78	4.55	–	–
t-stat	1.48	1.30	1.48	–	–

Table G.6. The heterogeneous intermediary mimicking portfolio (HIMP): Comparing models
This table presents time-series regression results of heterogeneous intermediary mimicking portfolio (HIMP) on mimicking portfolios for the representative intermediary factors in AEM and HKM according to: $\text{HIMP}_t = \alpha_{\text{MP}} + \beta'_{\text{FMP}}\,\text{FMP}_t + \epsilon_t$, where FMP is either the mimicking portfolio for broker-dealer leverage factor from AEM (AEM_MP), or the mimicking portfolio for capital factor for primary dealers' holding companies from HKM (HKM_MP), or both AEM_MP and HKM_MP. The factor-mimicking portfolios are constructed by projecting the heterogeneous intermediary, AEM's leverage, and HKM's capital factors unto the space of equity and bond returns according to equations (F.1) and (F.2). The sample is quarterly from 1970Q1 to 2017Q3. Standard errors are in parentheses.

	Dependent variable: HIMP				
	(1)	(2)	(3)	(4)	(5)
α_{MP}	5.03***	4.03***	4.14***	3.74***	3.91***
	(0.94)	(0.85)	(0.85)	(0.83)	(0.83)
AEM MP	0.72***			0.37***	0.49***
	(0.13)			(0.12)	(0.13)
HKM MP		0.94***	0.68**	0.82***	0.15
		(0.10)	(0.27)	(0.10)	(0.30)
MktRF			0.27		0.65**
			(0.26)		(0.27)
Observations	191	191	191	191	191
R^2	0.14	0.32	0.33	0.35	0.37
Adjusted R^2	0.14	0.32	0.32	0.35	0.36
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table G.7. The heterogeneous intermediary mimicking portfolio (HIMP): Comparing models with alternative projections

This table presents time-series regression results of heterogeneous intermediary mimicking portfolio (HIMP) on mimicking portfolios for the representative intermediary factors in AEM and HKM according to: $\text{HIMP}_t = \alpha_{\text{MP}} + \beta'_{\text{FMP}}\,\text{FMP}_t + \epsilon_t$, where FMP is either the mimicking portfolio for broker-dealer leverage factor from AEM (AEM_MP), or the mimicking portfolio for capital factor for primary dealers' holding companies from HKM (HKM_MP), or both AEM_MP and HKM_MP. The factor-mimicking portfolios are constructed by projecting the heterogeneous intermediary, AEM's leverage, and HKM's capital factors unto the space of equity and bond returns according to equations (F.1) and (F.2). The sample is quarterly from 1970Q1 to 2017Q3. Standard errors are in parentheses.

	Dependent variable: HIMP				
	(1)	(2)	(3)	(4)	(5)
α_{MP}	1.21***	0.96***	0.95***	0.88***	0.92***
	(0.25)	(0.22)	(0.23)	(0.22)	(0.22)
AEM MP	0.55***			0.23**	0.32***
	(0.09)			(0.09)	(0.11)
HKM MP		0.38***	0.39***	0.34***	0.13
		(0.04)	(0.11)	(0.04)	(0.14)
MktRF			−0.01		0.13
			(0.07)		(0.08)
Observations	191	191	191	191	191
R^2	0.16	0.35	0.35	0.38	0.38
Adjusted R^2	0.15	0.35	0.35	0.37	0.37
Note:			*p<0.1; **p<0.05; ***p<0.01		